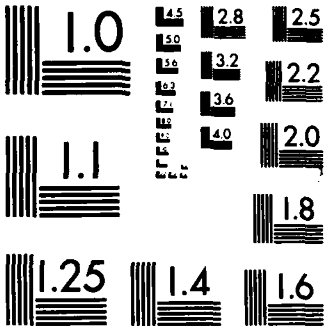
AD-A184 374   A COMPUTATIONAL EXAMINATION OF ORTHOGONAL DISTANCE   1/1
REGRESSION(U) COLORADO UNIV AT BOULDER DEPT OF COMPUTER
SCIENCE P T BOGGS ET AL 08 MAY 87 CU-CS-362-87
UNCLASSIFIED ARO-21453 10-MA DAAG29-84-K-0140   F/G 12/3   NL



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS-1963-A

UNIVERSITY OF COLORADO

A COMPUTATIONAL EXAMINATION OF ORTHOGONAL DISTANCE REGRESSION

Paul T. Boggs *
Janet R. Donaldson **
Robert B. Schnabel ***
Clifford H. Spiegelman #

CS-CU-362-87 May 1987

DEPARTMENT OF COMPUTER SCIENCE
CAMPUS BOX 430
UNIVERSITY OF COLORADO, BOULDER
BOULDER, COLORADO 80309-0430

Technical Report

2

A COMPUTATIONAL EXAMINATION OF ORTHOGONAL DISTANCE REGRESSION

Paul T. Boggs *
Janet R. Donaldson **
Robert B. Schnabel ***
Clifford H. Spiegelman #

CS-CU-362-87 May 1987

* Optimization Group/Scientific Computing Division, National Bureau of Standards, Gaithersburg, MD 20899

* Optimization Group/Scientific Computing Division, National Bureau of Standards, Boulder, CO 80303-3328

*** Department of Computer Science, University of Colorado, Boulder, Colorado 80309 and Optimization Group/Scientific Computing Division, National Bureau of Standards, Boulder CO 80303-3328 (Research supported by ARO contract DAAG 29-84-K-0140)

# Statistical Engineering Division, National Bureau of Standards, Gaithersburg, MD 20899 (Research supported in part under Office of Naval Research Office Contract N00014-86-F-0025)

THE FINDINGS IN THIS REPORT ARE NOT TO BE CONSTRUED AS AN OFFICIAL DEPARTMENT OF THE ARMY POSITION, UNLESS SO DESIGNATED BY OTHER AUTHORIZED DOCUMENTS.

Abstract

Classical or ordinary least squares (OLS) is one of the most commonly used criteria for fitting data to models and for estimating parameters. This is true even when a key assumption for its use, namely that the independent variables are known exactly, is violated. Orthogonal distance regression (ODR) extends least squares data fitting to problems with independent variables that are not known exactly. Theoretical analysis, however. shows OLS is preferable to ODR for *straight line functions* under certain conditions, even when there are measurement errors in the independent variable. This has lead some to conjecture that under some similar conditions OLS will also be preferable to ODR for *nonlinear functions* even though there are errors in the independent variable. Examined

~~In~~ this paper. we presents the results of ~~an~~ empirical study designed to examine whether ODR provides better results than OLS when there are errors in the independent variable. ~~We examine~~ a variety of functions, both linear and nonlinear, under a variety of experimental conditions. The results indicate that, for the data and performance criteria considered, ODR never performs appreciably worse than OLS and sometimes performs considerably better. This leads ~~us~~ to the conclusion that ODR *is* appropriate for a wide variety of practical problems.

keywords

errors in variables
Monte Carlo study
ordinary least squares
orthogonal distance regression

1 Introduction

Of all of the criteria used for fitting data to models or for estimating parameters, classical or ordinary least squares (OLS) is the one most commonly used. This continues to be the case even when the assumptions required to fully justify its use are not completely met. *Orthogonal distance regression*, or ODR, is designed to extend least squares data fitting to a class of problems which violate a key assumption for the use of OLS in parameter estimation, namely that the independent variables, x_i, are known exactly. In this paper, we compare the performance of OLS with that of ODR when this key assumption for OLS is violated.

The data fitting problem arises by considering a data set (x_i, y_i), $i = 1, \ldots, n$, that has been collected and a model that is purported to explain the relationship of $y_i \in \Re^1$ to $x_i \in \Re^m$. Specifically, if we assume there is no error in (x_i, y_i) and the true or actual value of the parameter vector $\beta^a \in \Re^p$ is known, then

$$y_i = f(x_i; \beta^a)$$

where f is a smooth function that can be either linear or nonlinear in x_i and β. Alternatively, if we suppose that the observations y_i contain actual, but unknown, additive errors $\epsilon_i^a \in \Re^1$, and that the observations x_i are known exactly, then y_i satisfies

$$y_i = f(x_i; \beta^a) - \epsilon_i^a \qquad i = 1, \ldots, n. \tag{1.1}$$

Finally, if we allow there to be additive errors in both x_i and y_i, then the data satisfy

$$y_i = f(x_i + \delta_i^a; \beta^a) - \epsilon_i^a \qquad i = 1, \ldots, n. \tag{1.2}$$

were $\delta_i^a \in \Re^m$ is the actual, but unknown, additive error in x_i. (Note that we have chosen the signs of ϵ_i and δ_i for convenience and consistency with other work.)

Using OLS, β^a is approximated by finding the β^{OLS} for which the sum of the squares of the n *vertical distances* from the curve $f(x_i; \beta)$ to the n data points is minimized. This is accomplished by the minimization problem

$$\min_{\beta} \sum_{i=1}^{n} w_i^2 (f(x_i; \beta) - y_i)^2 \tag{1.3}$$

where w_i, $i = 1,\dots,n$, are non-negative numbers that allow the procedure to be applied to problems when the observations should be weighted differently. When (1.1) is satisfied and $\epsilon = (\epsilon_1,\dots,\epsilon_n)^T \sim N(0,\sigma^2 I)$, then (1.3) with each $w_i = 1$ results in the maximum likelihood estimator of β^a.

Since (1.3) assigns all errors to y_i, *a critical assumption of OLS in parameter estimation is that there are no errors in x_i*. When this assumption is violated, use of OLS does not appear to be fully justified, and may not produce good estimates [Ful87, Mor71].

ODR, on the other hand, does allow for errors in x_i. ODR approximates β^a by finding that β for which the sum of the squares of the n *weighted orthogonal distances* from the curve $f(x_i;\beta)$ to the n data points is minimized. The estimated parameters, β^{ODR}, are then those values that solve the minimization problem

$$\min_{\beta,\delta} \sum_{i=1}^{n} w_i^2 \left[(f(x_i + \delta_i;\beta) - y_i)^2 + \delta_i^T \rho_i^2 \delta_i \right]. \qquad (1.4)$$

where $\rho_i \in \Re^{m\times m}$, $i = 1,\dots,n$, is a set of positive diagonal matrices that allow ϵ_i and δ_i to have different variances [BogBS85]. When (1.2) is satisfied and $\epsilon, \delta_1,\dots,\delta_n$ are independent and distributed as $\epsilon \sim N(0,\sigma_\epsilon^2 I)$ and $\delta_i \sim N(0,\sigma_\delta^2 \rho_i^{-1})$, then (1.4) with each $w_i = 1$ results in the maximum likelihood estimator of β^a. In the most common use of ODR, it is assumed that each $\rho_i = \rho I$, where ρ is the ratio of the standard deviations of the errors in the y and x data, i.e., $\rho = \sigma_\epsilon/\sigma_\delta$.

In this paper, we present the results of an empirical study designed to examine whether ODR provides better results than OLS when there are errors in both x_i and y_i. We examine a variety of functions, including functions nonlinear in x and β, under a variety of experimental conditions. While the statistical properties of the estimators from ODR fits are not yet well understood, [ReiGL86, Ful87] show that there are theoretical reasons to prefer OLS to ODR for a *straight line function* in certain situations even though there are errors in the observations x_i. This has led some to conjecture that under some similar conditions we should also ignore the errors in x_i when fitting models which are nonlinear in β or x. The results of our study indicate that this is probably not the case for the measures we have chosen. Specifically, for the data and performance criteria considered, ODR never performs appreciably worse than OLS and sometimes performs considerably better. This leads us to the conclusion that ODR is appropriate in a wide variety of practical problems.

To our knowledge, this is the first extensive computational study of the errors in variable problem with nonlinear functions. Previous work [e.g., Ful87, Mor71] has mainly concentrated on analytical analysis of the straight line function. We believe that this is partially due to the fact that until now, the ODR problem has been relatively expensive to solve and the necessary software has not been readily available. [BogBS85], however, presents a trust-region, Levenberg-Marquardt algorithm that exploits the structure of the ODR problem to obtain a procedure that is both stable and efficient. The order of operations per iteration, and the constant for the highest order term, are the same for the algorithm developed in [BogBS85] as for a trust region. Levenberg-Marquardt solution of the OLS problem, namely $O(np^2)$ operations per iteration. (A straight forward use of an OLS algorithm on (1.4) would require $O(n(n+p)^2)$ operations per iteration [BogBS85] which is clearly prohibitive for large values of n.) The algorithm described in [BogBS85] has been implemented in the portable Fortran subroutine library ODRPACK [BogBDS87]. The availability of ODRPACK makes it reasonable to conduct the study reported here.

We emphasize that this study is only a first step and that we have left many important questions unanswered. Some of these are discussed in §2 where we detail the motivation for our study and its scope. We outline our Monte Carlo procedure in §3, and in §4 we summarize our observations and present our conclusions. Our plans for future work are given in §5. A detailed description of our results and the accompanying figures are presented in the Appendix.

2 Motivation

While ODRPACK provides an efficient means of solving ODR, it is not known whether there is a "theoretical" penalty for using ODR since the theoretical analysis of ODR is not yet available for functions other than a straight line.

For a straight line, [ReiGL86] notes that OLS results in a smaller mean square error of the slope than ODR when

$$\left(\frac{B}{\rho}\right)^2 < \frac{2}{n-2}$$

where B is the slope of the line, $\rho = \sigma_\epsilon/\sigma_\delta$, and it is assumed that ϵ_i and δ_i are independent. [BogBS85], on the other hand, presents empirical results for which ODR appears preferable to OLS. We are therefore interested in studying the question

> *Under what conditions is ODR preferable to OLS, and, conversely, when is OLS preferable to ODR?*

This paper is a first approximation to answering this question.

The question actually has two parts. First, can we detect a practical difference in performance between ODR and OLS? Second, assuming that differences in performance are detected, can we characterize the conditions under which such differences occur in order to predict when one method will be preferable to the other?

To detect a difference between ODR and OLS we must select a measure of performance. As a first step, we have chosen to investigate the estimated parameter values, $\hat{\beta}$, and function values, $f(x_i;\hat{\beta})$, since these are commonly of interest and easily understood. For both we use three standard measures of performance: bias, variance and mean square error.

To determine performance predictors that can be used to characterize *a priori* whether a data set should be solved using ODR or OLS, we reexamine the results for a straight line function. The ODR solution for a straight line can be derived by noting that the square of the weighted orthogonal distance between the line $\beta_1 x + \beta_2$ and the data point (x_i, y_i) is

$$\frac{(\beta_1 x_i + \beta_2 - y_i)^2}{1 + \left(\frac{\beta_1}{\rho}\right)^2}.$$

If we assume that any function, whether linear or nonlinear in x and β, is at least approximately a straight line in the neighborhood about each individual point (x_i, y_i), then the square of the weighted orthogonal distance between $f(x;\beta)$ and (x_i, y_i) is

$$g_i(\beta) = \frac{(f(x_i;\beta) - y_i)^2}{1 + \left(\frac{\partial f(x_i;\beta)/\partial x}{\rho}\right)^2}$$

meaning that the ODR problem (1.4) can be approximated by

$$\min_{\beta_i} \sum_{i=1}^{n} w_i^2 g_i(\beta). \tag{2.1}$$

As the ratios

$$h(x_i; \beta) = \frac{\partial f(x_i; \beta)/\partial x}{\rho}$$

approach 0. (2.1) becomes equivalent to the OLS problem (1.3). The sizes of the ratios $h(x_i; \beta)$ should thus be related to the question of when ODR is different from OLS.

Unfortunately. when $f(x_i; \beta)$ is not linear in x_i. $\partial f(x_i; \beta)/\partial x$ and therefore $h(x_i; \beta)$ varies with x_i. Consequently. in order to assess a single number as a performance predictor of ODR in relation to OLS we must map $\partial f(x_1; \beta)/\partial x, \ldots, \partial f(x_n; \beta)/\partial x$ into a single value. For simplicity. we initially choose

$$Q = \max\{|\partial f(x_i; \beta)/\partial x|,\ i = 1, \ldots, n\},$$

i.e.. the ℓ_∞ norm of $[\partial f(x_1; \beta)/\partial x, \ldots, \partial f(x_n; \beta)/\partial x]^{\mathrm{T}}$. as the mapping and

$$Q/\rho$$

as the performance predictor. We note that other norms could be chosen.

Our approach for this initial pilot study. described in detail in §3, is relatively simple. Briefly, we select seven functions that, although clearly not exhaustive, are ubiquitous in science and engineering. and seven different values of ρ that are used with each function. For each function, we also choose two parameter sets that produce different values of Q. Treating ρ as known. we then examine (a) how the performance of ODR and OLS varies for an individual function as ρ changes, (b) how the results vary between functions and (c) how well the performance prediction value Q/ρ forecasts the observed results.

We recognize that we are examining our data under ideal conditions. Clearly. ρ and Q will frequently not be known exactly. but this does not seem to be bothersome. In our experience. Q can be reasonably estimated for most functions and data sets. Furthermore. if $\hat{\beta}(\rho)$ is the estimate of β^a for a given value of ρ. we can show that when δ is small, then $d\hat{\beta}(\rho)/d\rho$

is also small. Thus, the value of $\hat{\beta}$ should not change much as ρ is varied. This is observed in other experience not reported in this paper. For the remainder of this paper, therefore, we assume that the true value of ρ is known, but see §5.

3 Procedure

In this section, we briefly describe the details of our Monte Carlo study.

Our study examines seven functional forms.

$$
\begin{aligned}
f_1(x_i;\beta) &= \beta_1 x + \beta_2 && (3.1)\\
f_2(x_i;\beta) &= \beta_1 x^2 + \beta_2 && (3.2)\\
f_3(x_i;\beta) &= \beta_1 x^2 + \beta_2 x + \beta_3 && (3.3)\\
f_4(x_i;\beta) &= \beta_1 \exp(\beta_2 x) && (3.4)\\
f_5(x_i;\beta) &= \beta_1 \exp(\beta_2 x) + \beta_3 && (3.5)\\
f_6(x_i;\beta) &= \beta_1 \sin(\beta_2 x + 2) && (3.6)\\
f_7(x_i;\beta) &= \beta_1 \sin(\beta_2 x + \beta_3) && (3.7)
\end{aligned}
$$

We have selected two sets of parameter values for each function. For both,

$$\max\{|y_i^a| = |f(x_i;\beta^a)|,\ i = 1,\ldots,n\} \approx 1.$$

In addition, the first parameter set is chosen so that $Q \approx 1$ and the second so that $Q \approx 10$. The data sets constructed for each function using the first parameter set thus have some similar attributes, as do those for the data sets constructed using the second parameter set, even though the functional forms are different. The parameter sets are as follows.

Function	Parameter Set 1	Parameter Set 2
f_1	$\beta^a = (1.1,\ 0.9)^T$	$\beta^a = (10.0,\ -2.0)^T$
f_2	$\beta^a = (0.3,\ 3.3)^T$	$\beta^a = (4.5,\ -3.0)^T$
f_3	$\beta^a = (0.3,\ 0.4,\ 3.3)^T$	$\beta^a = (4.5,\ 1.0,\ -4.5)^T$
f_4	$\beta^a = (0.4,\ 1.0)^T$	$\beta^a = (1.2,\ 1.6)^T$
f_5	$\beta^a = (0.4,\ 1.0,\ -1.0)^T$	$\beta^a = (1.2,\ 1.6,\ -5.0)^T$
f_6	$\beta^a = (0.9,\ 1.1)^T$	$\beta^a = (5.0,\ 2.0)^T$
f_7	$\beta^a = (0.9,\ 1.1,\ 2.0)^T$	$\beta^a = (5.0,\ 2.0,\ 1.0)^T$

For all functions and parameter sets, the number of observations, n, is 51, and $x_i^a, i = 1, \ldots, n$ are the 51 equally spaced values over the interval $[-1, 1]$. The ranges of x_i^a and y_i^a are thus comparable.

We analyze each of the seven functions using both sets of parameters and seven different values of $\rho = \sigma_\epsilon / \sigma_\delta$, namely $\rho = \frac{1}{10}, \frac{1}{2}, 1, 2, 10, 100$, and ∞. When $\rho \neq \infty$, we analyze the data using both methods, ODR and OLS. (When $\rho = \infty$ the two methods are equivalent.) A total of 182 combinations of function, parameter set, ρ, and method are considered.

We generate two groups of 500 data sets each for this collection of 182 combinations. In the first group, a *single* set of values $\tilde{\epsilon}$ and $\tilde{\delta}$ is used to produce the actual errors ϵ_i^a and δ_i^a for each of the problems within the collection. The problems analyzed within the collection of 182 combinations in the first group, therefore, are *not* independent. In the second group, a different set of errors $\tilde{\epsilon}$ and $\tilde{\delta}$ are generated for each of the problems within the collection. The problems analyzed within the collection of 182 combinations in the second group *are* independent. The first group allows us to make pair-wise comparisons of the individual results obtained using ODR and OLS, and enables us to ensure that variations in performance are not artificially induced by variation in the data used within the collection. The independent data used for the second group open our analysis to a wider range of statistical tests.

For each of the 500 data sets in both groups, we generate the n i.i.d. pseudo random values, $\tilde{\epsilon}_i$, $i = 1, \ldots, n$, from a normal distribution with mean 0 and standard deviation .05, and the n values $\tilde{\delta}_i$, $i = 1, \ldots, n$, also i.i.d. normally distributed with mean 0 and standard deviation .05. This set of values $\tilde{\epsilon}$ and $\tilde{\delta}$ is used to produce the actual errors ϵ_i^a and δ_i^a, where

$$\epsilon_i^a(\rho) = \rho \left(\frac{2}{\rho^2 + 1} \right)^{1/2} \tilde{\epsilon}_i$$

and

$$\delta_i^a(\rho) = \left(\frac{2}{\rho^2 + 1} \right)^{1/2} \tilde{\delta}_i$$

for each of the seven values of ρ. (The expected sum of the squared errors is thus constant over the seven different values of ρ.) These errors $\epsilon_i^a(\rho)$ and $\delta_i^a(\rho)$ are then used to generate the "observations" $y_i(\rho)$ and $x_i(\rho)$ for a given value of ρ, where

$$y_i(\rho) = y_i^a - \epsilon_i^a(\rho) = f(x_i^a; \beta^a) - \epsilon_i^a(\rho)$$

and

$$x_i = x_i^a + \delta_i^a(\rho).$$

The errors are produced using the Marsaglia and Tsang [MarT84] pseudonormal random number algorithm as implemented by James Blue and David Kahaner of the Scientific Computing Division of the National Bureau of Standards. The ODR and OLS estimators are computed using ODRPACK. Parameters are initialized to β^a for both the ODR and OLS solutions, and for the ODR solutions the errors in x_i are initialized to δ_i^a and ρ is set to the correct value. Using β^a and δ_i^a for starting values is reasonable since in this study we are only interested in the properties of the ODR and OLS solutions and not in the properties of the estimation procedures used to obtain them. The graphics package TEMPLATE [Meg86] is used to produce the plots.

All computations are performed in single precision on the CDC Cyber 205 at the National Bureau of Standards. Approximately 3400 seconds cpu time are required to solve the 182000 optimization problems. There are 31 trials for which one of the 182 problems failed to converge in 200 iterations. Each of these "failed" trails is omitted from the analysis, and another trial substituted in its place. Such a small percentage of failures does not affect our conclusions.

4 Conclusions

Our study addresses two main issues:

1. the relationship between the performance of ODR and OLS for parameter and function estimation as determined by the three measures, bias, variance and mean square error (mse); and,

2. to a lesser extent, how the performance characteristics of ODR and OLS vary for different values of Q and ρ.

The first item is important for determining whether there is a preferred method. The second may be important in determining how to choose between the two methods. In this section, we summarize our observations and conclusions regarding these two issues. A more detailed description of the results from our study is given in the Appendix.

Our results indicate that ODR should always be used when our criteria are relevant. A subroutine library such as ODRPACK is just as easy to use as an OLS subroutine library and is no more computationally expensive per iteration. More importantly, however, for all of the measures of performance examined in our study. ODR is seldom seriously worse than OLS and is frequently significantly better. especially for $\rho \leq 2$. We conclude that. except for outliers, ODR results in smaller bias. variance. and mse for both parameter and function estimates than does OLS.

- Our results for the bias of the parameter and function estimates are especially clear.
    - OLS is statistically better (as described in the Appendix) only 2% of the time.
    - ODR. on the other hand, is appreciably better more than 50% of the time, and the largest of the relative differences between the ODR and OLS biases when the ODR bias is closer to 0 is more than 250 times the largest of the relative differences observed when the OLS bias is closer to 0.
- Our results for the variance of the parameter and function estimates also decisively favor ODR over OLS.
    - For both the variances of the parameter estimates and the variances of the function estimates, the ODR variance is appreciably smaller than the OLS variance more than 23% of the time. and in over 10% of the cases the ODR variance is less than 50% of the the OLS variance.
    - Conversely. OLS results in appreciably smaller variance than ODR only 2% of the time.

The 2% of the time that OLS has appreciably smaller variance than ODR all occur for the same two data sets involving the sine function (3.6) and (3.7) when $\rho = \frac{1}{10}$. The results for these two data sets, shown in figures 4 and 5, are clearly different from any of the other data sets we examined in that they contain significant outliers. Further analysis will be required to explain their anomalous behavior.

- Our results for the mse of the parameter and function estimates are essentially the same as those observed for the variance:

  - ODR results in appreciably smaller mse approximately 25% of the time with the ODR mse less than 50% of the OLS mse approximately 20% of the time, while
  - OLS results in appreciably smaller mse only 2% of the time.

  Again. the times OLS has smaller mse are all observed for the two data sets that affected the variance in the analogous manner and will require further study.

The 500 data sets used in our study is apparently not enough to confirm the theoretical results reported in [ReiGL86] that indicate OLS should be preferable under certain conditions when the function is a straight line. OLS does. in fact, produce a smaller variance and mse than ODR for our linear data sets. One would seldom, if ever, call the difference statistically or practically significant, however. This. coupled with the bias data that indicates that ODR *is* significantly better for a straight line, leads us to conclude that ODR is the method of choice for our criteria.

Our final conclusion. based on a visual examination of the almost 200 plots generated for this study. is that Q/ρ does not adequately predict the relationship between the performance of ODR and OLS, although as a crude measure it does have its merits. Since our primary interest originally was to predict when one should prefer ODR to OLS, or visa versa. and since we conclude now that we always prefer ODR to OLS when there are errors in x and ρ is known. the failure of this performance predictor is not important to our current results. Further analysis of the predictors mentioned in §2 may be required. however. when we examine other performance criteria, or when we examine the effect of not knowing ρ exactly.

5 Future Work

As we have mentioned before. this is a pilot study. and as one would expect from such a study. we have answered some questions and raised others.

There are clearly additional questions to be answered from our current data. The anomalous results we noted for functions (3.6) and (3.7) definitely require further analysis. In addition. we would like to examine other measures of performance. and the structure of the estimated residuals. Further analysis may also raise the need for different values to help predict whether to use ODR or OLS as described in §2.

Additional studies need to be performed. Our current plans include an examination of the effect of using an incorrect value of ρ. We also plan to examine other functions with much steeper slopes than what we allowed in this study. and with unequally spaced x data. Finally. we note that least squares estimates. whether OLS or ODR. are not robust in the presence of outliers [Ful87]. and that [GolV83] shows that ODR problems are more ill conditioned that the corresponding OLS problem. We would therefore like to experiment with diagnostics and resampling techniques such as bootstrapping that could be used to indicate when the ODR results might be affected by ill conditioning or outliers.

Acknowledgements. The authors wish to thank R. Carroll (University of North Carolina) and R. Byrd (University of Colorado at Boulder) for many useful discussions and helpful comments concerning this study.

A Results and Observations

In this section, we present a detailed description of the results and observations that support the conclusions presented in §4.

Because of the large amount of data examined, our analysis is primarily graphical. We have included in this paper only representative examples of the almost 200 plots generated for this study. As noted in §3, our study includes two groups of 500 data sets: the first with errors ϵ^a and δ^a within the collection of 182 combinations of function, parameter set, ρ, and method that are *not* independent; and the second with errors ϵ^a and δ^a within the collection of 182 combinations that *are* independent. All plots shown here are from the first group. The text describes additional observations derived from the second group. The full graphical analysis of both groups is available in [BogDSS87].

A.1 Parameter Estimates

A.1.1 Bias

Results. To determine how close our estimated parameter values, $\hat{\beta}$, come to the actual or true values, β^a, we examine the bias of the estimated parameters, i.e., the values

$$\hat{\beta}_j - \beta_j^a$$

where $\hat{\beta}_j$ designates the estimated value of the j-th parameter for a given problem and data set, and β_j^a is the corresponding true value of the parameter.

For each of the two groups of data sets, we display the 500 parameter biases obtained for each of the parameters in the collection of 182 combinations of function, parameter set, ρ, and method. We then examine each of the resultant pairs of bias estimates obtained using the two methods.

Figures 1, 2 and 3 show these results for each of the parameters and values of ρ for three representative combinations of function and parameter set. These three figures correspond to three functions of increasing complexity. Figure 1 shows function (3.1) parameter set 2, a straight line function with slope of 10. Figure 2 shows function (3.3) parameter set 1, a quadratic function in x with maximum slope 1 for $x \in [-1,1]$. Figure 3 shows the results for function (3.5) parameter set 2, an exponential function that is nonlinear in both x and β and that has maximum slope 10 for $x \in [-1, 1]$.

Figures 4 and 5 show the bias results for the sine functions (3.6) and (3.7), respectively, both using parameter set 2, and are analogous to figures 1, 2 and 3. These two figures are not typical in that they show a small number of outliers when $\rho = \frac{1}{10}$; no outliers are observed in *any* of the other parameter bias plots. These outliers appear to adversely affect the corresponding variance and mean square error estimates. as discussed in the following sections. as well as the results for the function estimates for these data sets.

Each column of icons in these figures represents a modified box-and-whisker plot [Tuk77]:

$\circ$ designates the median.

$+$ designates the quartiles. and

$\diamond$ designates the maximum and minimum.

The remaining bias values from each of the 500 data sets are designated by a dot ($\cdot$). The values are grouped by ρ, then by parameter and finally by method. Thus, the first column of icons on each of these figures displays the bias observed using ODR for β_1 when $\rho = \frac{1}{10}$, the second column displays the bias using OLS for β_1 when $\rho = \frac{1}{10}$. etc.

If the median parameter bias is determined to be different from 0 at the .05 significance level using a two-sided sign test, the median is "flagged" with a check ($\surd$) plotted above the corresponding icon for the maximum value. (See, e.g., figure 1, $\rho = \frac{1}{10}$, β_1 estimated using OLS.) If it is not different at the .05 significance level, no flag is shown.

Observations. The parameter bias results are essentially the same for both groups of data.

- In more than 33% of the ODR/OLS pairs. the sign test indicates that the median ODR bias is 0 (i.e., that the hypothesis that the median ODR bias is equal to 0 would not be rejected at the .05 significance level) when the median OLS bias is not 0.

- In fewer than 2% of the ODR/OLS pairs, the sign test indicates the ODR median is not 0 when the OLS median is 0.
- In approximately 25% of the ODR/OLS pairs, the sign test indicates that both medians are not 0. Of these cases, when it is possible to visually detect a difference in the medians of these pairs, the median ODR bias almost always closer to zero.

For the largest of the differences between the median ODR and OLS biases, the median ODR bias is always closer to 0 than the median OLS bias. The largest differences occur for values of $\rho \leq 1$, with the differences increasing as ρ decreases. As noted in the next section, the variance of the ODR results is generally as small or smaller than that of the OLS results.

A.1.2 Sample Variance

Results. The sample variance, $\hat{\sigma}^2_{\hat{\beta}_j}$, of the parameter estimate, $\hat{\beta}_j$, is a measure of the variability of the estimate about its average value. To examine the relationship between the variance of the ODR parameter estimates and variance of the OLS parameter estimates, we plot the base 10 logarithm of the ratios of the sample variances for each of the estimated parameters, i.e.,

$$\log\left[\frac{\hat{\sigma}^2_{\hat{\beta}_j^{ODR}}}{\hat{\sigma}^2_{\hat{\beta}_j^{OLS}}}\right] = \log\left[\frac{\sum_{i=1}^{500}\left(\hat{\beta}_j^{ODR} - \frac{\sum_{i=1}^{500}\hat{\beta}_j^{ODR}}{500}\right)^2}{\sum_{i=1}^{500}\left(\hat{\beta}_j^{OLS} - \frac{\sum_{i=1}^{500}\hat{\beta}_j^{OLS}}{500}\right)^2}\right]$$

against

$$-\log\left[\frac{Q}{\rho}\right].$$

These plots allow us to examine the relationship between the individual variance pairs as well as how the relationship changes as a function of Q and ρ. All resultant variance ratios are examined.

Figures 6, 7 and 8 are representative examples of the variance plots, and show the variance ratios for the data shown in figures 1, 2 and 3, respectively. The icon used for each ratio is the number of the subscript of

β. i.e., the ratio of the variances of β_1 are plotted using the symbol "1," the ratio of the variances of β_2 are plotted using the symbol "2," etc. Note that the columns of icons in these figures are in the same order with respect to ρ as the "Parameter Bias" plots.

For the first group of data, the errors, and therefore the variances, are not independent. We thus use a two-sided Pitman Nearness Test [Rao86] to make a pair-wise comparison of the 500 deviations

$$\left|\hat{\beta}_j - \text{median}\{\hat{\beta}_j\}\right|$$

obtained for each parameter of each function and parameter set using each of the values of ρ. Note that we expect the dependence introduced by the sample median to be small. If, using this test, we reject the hypothesis that the deviations from the two methods are the same at the .05 significance lever, we "flag" the appropriate icon with an asterisk ($*$). On these variance plots, we also indicate the magnitude of the ratios with the two lines marked "20 PERCENT." An icon falling above the upper of these two lines indicates the ODR variance is more than 20% bigger than the OLS variance. An icon falling below the lower of these two lines indicates the OLS variance is more than 20% bigger than the ODR variance.

For the second group of data, the errors, and therefore the variances, are independent. We are thus able to test whether each variance ratio is different from 1 at the .05 significance level using a two-sided F-test.

Observations. In the first group of data, the two-sided Pitman Nearness Test indicates the deviations from the ODR fit are different than those obtained from the OLS fit 101 times out of the resultant 204 pairs.

- In 83 of the 101 cases, the ODR variance is smaller than the OLS variance. These 83 cases include all 62 cases where the OLS variance is more than 20% larger than the ODR variance, and the 41 times that the OLS variance is more than twice the OLS variance.
- In 18 of the 101 cases, the OLS variance is smaller than the ODR variance. These 18 cases include only 2 of the 5 times that the ODR variance is more than 20% larger than the OLS variance, and they include neither of the 2 times that the ODR variance is more than

twice the OLS variance. (Each of the 5 cases where the ODR variance is more than 20% larger than the OLS variance occur for the sine functions (3.6) and (3.7) using parameter set 2 at $\rho = \frac{1}{10}$.)

For the second group of data. the F-test indicates that the variances using ODR and OLS are different at the .05 significance level 69 times.

- In 64 of these 69 cases, the ODR variance is significantly smaller than the OLS variance. and in 36 of the 64 cases the ODR variance is less than 50% of the OLS variance.

- In 5 of these 69 cases. the OLS variance is significantly smaller than the ODR variance. and the OLS variance is less than 50% of the ODR variance in 2 of these 5. Both of these 2 cases occurred in the sine functions results using parameter set 2 at $\rho = \frac{1}{10}$. The bias results for these two functions. like the results shown in figures 4 and 5 for the first group of data. include a small number of outliers which appear to be responsible for the increased ODR variance. Such outliers were not observed in *any* of the other results. including functions (3.6) and (3.7) using parameter set 2 when $\rho \geq \frac{1}{2}$.

For both the first and second data groups. the largest differences between the variances occur for the smaller values of ρ. and, with the exception of the sine function data using parameter set 2 at $\rho = \frac{1}{10}$. when the difference between the ODR and OLS variances are large, the ODR variance is the smaller of the pair.

A.1.3 Mean Square Error

Results. The mean square errors (mse) of the estimated parameters are measures of the variability of the estimate β_j about its true value. β_j^a. We examine the relationship between the mse observed using ODR and OLS in the same manner that we analyze the sample variance of the parameter estimates. We plot the base 10 logarithm of the ratios of the mse of the estimated parameters. i.e.,

$$\log \left[\frac{\sum_{i=1}^{500} \left(\hat{\beta}_j^{ODR} - \beta_j^a \right)^2}{\sum_{i=1}^{500} \left(\hat{\beta}_j^{OLS} - \beta_j^a \right)^2} \right]$$

against

$$-\log\left[\frac{Q}{\rho}\right].$$

All resultant mse ratios are examined.

Figures 9, 10 and 11, corresponding to the data shown in figures 1, 2 and 3. respectively, are representative examples of these plots. Like the sample variance plots discussed in §A.1.2, the icon used for each ratio is the number of the subscript of β. For both groups of data, we indicate the magnitude of the ratios with the two lines marked "20 PERCENT." For the first group of data, we also use a two-sided Pitman Nearness Test to make a comparison of the 500 values of $\left|\hat{\beta}_j - \beta_j^a\right|$ observed for each of the parameters of each of the functions at each of the values of ρ. If, using this test, we reject the hypothesis that the magnitudes of the biases for the two methods are not the same at the .05 significance lever, we "flag" the appropriate icon with an asterisk (*).

Observations. For the first group of data, with errors that are not independent, the Pitman Nearness Test indicates that the deviations from the ODR fit are different than those obtained from the OLS fit 119 times out of the 204 resultant ODR/OLS pairs.

- In 100 of the 119 cases, the ODR mse is smaller than the OLS mse. These 100 include most of the 91 ratios for which the OLS variance is more than 20% larger than the OLS variance, and all of the 53 cases for which the observed ODR mse is less than 50% of the OLS mse.

- In 19 of the 119 cases, the OLS mse is smaller than the ODR mse, including all 3 cases (each resulting from the sine functions with parameter set 2 at $\rho = \frac{1}{10}$) for which the ODR variance is more than 20% larger than the OLS variance.

For the second group of data, with errors that are independent, we have not performed any statistical test of significance.

- We observe, however, that for 90 of the 204 resultant ratios the OLS mse is more than 20% bigger than the ODR mse, and in 54 of these the OLS mse is more than twice the ODR mse.

- Also, the ODR mse is more than 20% larger than the OLS mse in only 4 of the 204 resultant ratios, and the ODR mse is more than twice the OLS mse in only 2 of these. (Both of the two ratios for which the ODR mse is twice the OLS mse again result from the sine functions using parameter set 2 for $\rho = \frac{1}{10}$.)

Like the variance results, for both groups of data the largest differences occur for the smaller values of ρ. Also, for the largest differences, the ODR mse is almost always the smaller of the two mse.

A.2 Function Estimates

A.2.1 Bias

Results To determine how close the function estimates, $f(x_i;\hat{\beta})$, come to the actual or true values, $f(x_i;\beta^a)$, we examine the biases of the function estimates, i.e., the values

$$f(x_i;\hat{\beta}) - f(x_i;\beta^a)$$

where $\hat{\beta}$ designates the estimated value of the parameters for a given problem and data set. These are computed for 3 representative values of x_i over the interval [-1,1], namely, $x = -1, 0$ and 1. The data for each of the two groups of data sets are examined using modified box-and-whisker plots as were the parameter bias data. All resultant pairs of function estimates are examined.

Figures 12, 13 and 14 show representative examples of these plots, and present results for the same data sets as those analyzed in figures 1, 2 and 3. These figures are completely analogous to the parameter bias plots discussed in §A.1.2. Again, we test whether the median bias of the function estimate is different from 0 at the .05 significance level using a two-sided sign test, indicating medians that are not 0 according to this test with a check ($\surd$).

Observations. The bias results for the function estimates are almost exactly the same for both data groups.

- In more than 40% of the ODR/OLS pairs, the sign test indicates that the median ODR function bias is 0 while the median OLS function bias is not 0.

- In fewer than 2% of the ODR/OLS pairs, the sign test indicates that the median OLS function bias is 0 while the median ODR function bias is not 0.

- In the approximately 20% of the ODR/OLS pairs that both medians are not 0. the ODR median is almost always closer to 0 than the OLS median. and sometimes appreciably so.

For the largest of the differences between the medians. the median ODR function bias is always closer to 0 than the corresponding median OLS function bias. As was true for the parameter bias results. the largest differences occur for values of $\rho \leq 1$, with the differences increasing as ρ decreases. Also. as noted in the next section. the variance of the ODR results are generally as small or smaller than the corresponding OLS variance.

A.2.2 Sample Variance

Results. The sample variance, $\hat{\sigma}^2_{f(x_i;\hat{\beta})}$, is a measure of the variability of the function estimate $f(x_i;\hat{\beta})$ about its average value. To examine the relationship between the variance of the ODR function estimates and the variance of the OLS function estimates, we plot

$$\log\left[\frac{\hat{\sigma}^2_{f(x_i;\beta^{ODR})}}{\hat{\sigma}^2_{f(x_i;\beta^{OLS})}}\right] = \log\left[\frac{\sum_{i=1}^{500}\left(f(x_i;\hat{\beta}^{ODR}) - \frac{\sum_{i=1}^{500} f(x_i;\hat{\beta}^{ODR})}{500}\right)^2}{\sum_{i=1}^{500}\left(f(x_i;\hat{\beta}^{OLS}) - \frac{\sum_{i=1}^{500} f(x_i;\hat{\beta}^{OLS})}{500}\right)^2}\right].$$

against

$$-\log\left[\frac{Q}{\rho}\right].$$

for each of the function estimates observed at the three selected values of x_i. All resultant variance ratios are examined.

Figures 15, 16 and 17 are representative examples of these variance plots. and show the variance ratios corresponding to the function bias data shown in figures 12, 13 and 14. respectively. The format is analogous to that used for the parameter variance plots. discussed in §A.1.2. For the first data group, in which the errors are not independent. we test whether the deviations

$$\left|f(x_i;\hat{\beta}^{ODR}) - \text{median}\left\{f(x_i;\hat{\beta}^{ODR})\right\}\right|$$

are different from

$$\left|f(x_i;\hat{\beta}^{OLS}) - \text{median}\left\{f(x_i;\hat{\beta}^{OLS})\right\}\right|$$

at a .05 significance level using a two-sided Pitman Nearness Test, "flagging" with an asterisk (*) the ratios for which the absolute values of the deviations are found to differ at this significance level. For the second group. in which the errors are independent. we test whether each variance ratio is different from 1 at the .05 significance level using a two-sided F-test.

Observations. In the first group of data. the above mentioned two-sided Pitman Nearness Test indicates that the deviations obtained using ODR are different from those obtained using OLS at the .05 significance level 117 times out of 252.

- In 84 of the 117 cases. the ODR variance is smaller than the OLS variance. These 84 cases include 59 of the 68 cases where the OLS variance is more than 20% larger than the ODR variance, and all of the 32 cases where the OLS variance is more than twice the ODR variance.

- In 33 of the 117 cases. the OLS variance is smaller than the ODR variance. These 33 cases include 4 of the 5 cases that the ODR variance is more than 20% larger than the OLS variance. and 3 of the 4 cases where the ODR variance is more than twice the OLS variance. (The 5 cases that the ODR variance is more than 20% larger than the OLS variance all occur for the sine functions using parameter set 2 when $\rho = \frac{1}{10}$.)

In the second group of data, the F-test indicates that the variances using ODR and OLS are different at the .05 significance level 71 times out of the 252 resultant ra.ios.

- In 63 of the 71 cases, the ODR variance is significantly smaller than the OLS variance. and in 28 of these 63 cases the ODR variance is less than 50% of the OLS variance.

- In 8 of the 71 cases. the OLS variance is significantly smaller than the ODR variance. and the OLS variance is less than 50% of the ODR variance in 2 of these 8. Both of these 2 again result from the sine functions using parameter set 2 at $\rho = \frac{1}{10}$.

For both groups. the largest differences between the variances occur for the smaller values of ρ. and. with the exception of the sine function results for parameter set 2 when $\rho = \frac{1}{10}$. when the differences between the ODR and OLS variances are large. the ODR variance is the smaller of the two.

A.2.3 Mean Square Error

Results. The mse of the function estimate is a measure of the variability of the estimate $f(x_i; \hat{\beta})$ about its true value. $f(x_i; \beta^a)$. We examine the relationship between the mse of the function estimate observed using ODR and OLS in the same manner that we analyze the mse of the parameter estimates. namely. we plot

$$\log \left[\frac{\sum_{i=1}^{500} \left(f(x_i; \hat{\beta}^{ODR}) - f(x_i; \beta^a) \right)^2}{\sum_{i=1}^{500} \left(f(x_i; \hat{\beta}^{OLS}) - f(x_i; \beta^a) \right)^2} \right]$$

against

$$-\log \left[\frac{Q}{\rho} \right].$$

All resultant mse ratios are examined.

Figures 18. 19 and 20. corresponding to the data shown in figures 12. 13 and 13. respectively. are representative examples of these plots. The format for these plots. and the analysis performed. is analogous to that described in §A.1.3.

Observations. For the first group of data, with errors that are not independent. the Pitman Nearness Test indicates that the deviations from the ODR fit are different than those obtained from the OLS fit 157 out of the 252 resultant ODR/OLS pairs.

- In 132 of the 157 cases. the ODR mse is smaller than the OLS mse. These 132 include all of the 111 ratios when the OLS mse is more than 20% larger than the ODR mse. and all of the 64 ratios when the OLS mse is more than twice the ODR mse.

- In 25 of the 157 cases. the OLS mse is smaller than the ODR mse. including all 4 ratios for which the ODR mse is more than 20% of the OLS mse. and the 3 ratios for which the ODR mse is more than twice the OLS mse. (All 4 of these cases again occur for the sine functions using parameter set 2 when $\rho = \frac{1}{10}$.)

For the second group of data. with errors that are independent. we have not performed any statistical test of significance.

- We observe. however. that in 112 of the resultant 252 ratios. the OLS mse is more than 20% larger than the corresponding ODR mse. and in 59 cases the OLS mse is more than twice the corresponding ODR mse.

- Also. we observe that the ODR mse is more than 20% larger than the OLS mse in only 3 cases. and is more than twice the corresponding OLS mse in only 2 cases.

The mse results are thus similar to those we observed for the other performance measures in that both groups of data show that the differences in the mse increase as ρ decreases. Again. for the larger differences, with the exception of the outlier data in the sine function results. the ODR mse is always smaller than the OLS mse.

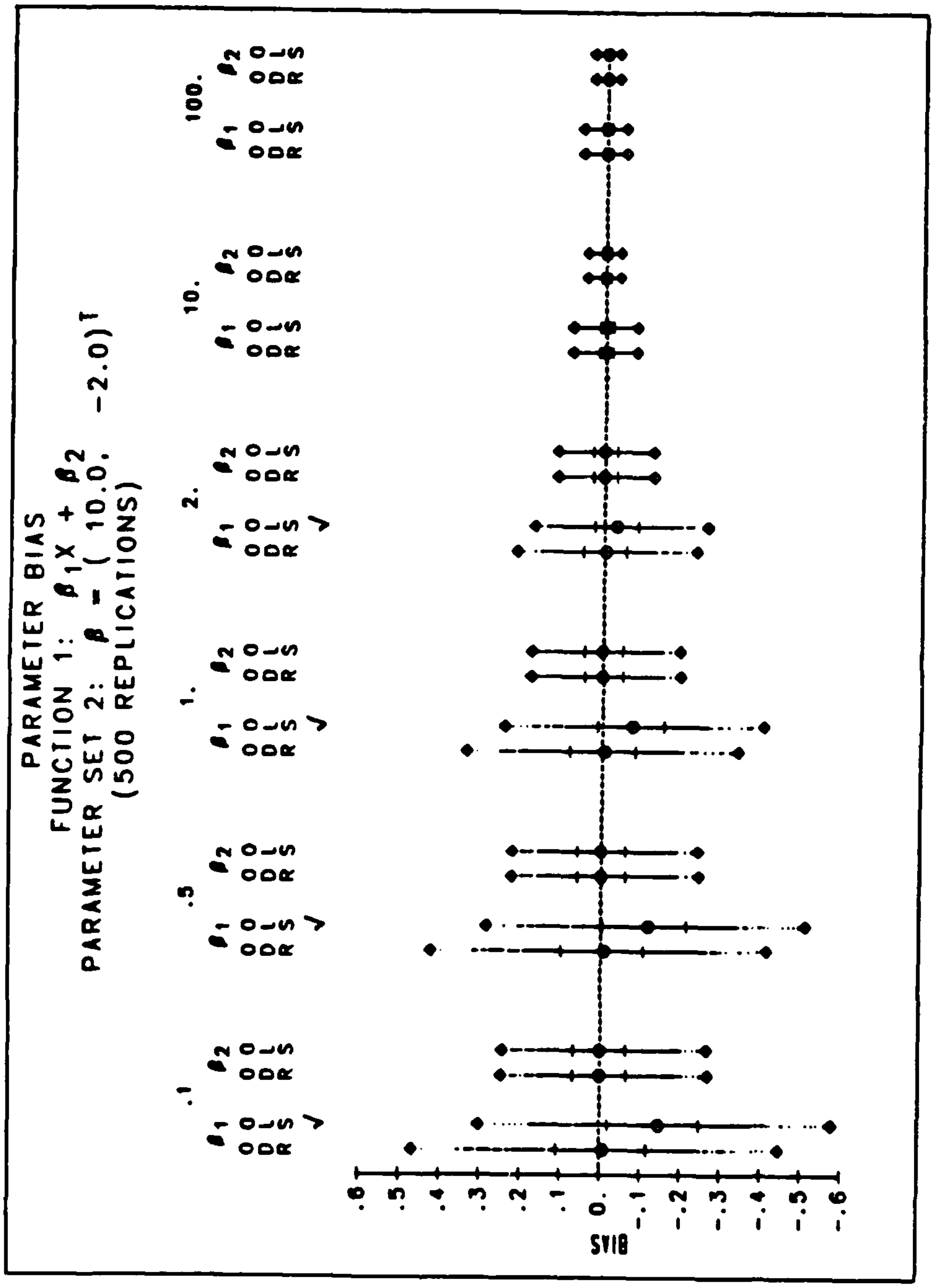

PARAMETER BIAS
FUNCTION 1: β1X + β2
PARAMETER SET 2: β = (10.0, -2.0)T
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
β1
β2
ODR
OLS
BIAS
.6
.5
.4
.3
.2
.1
0.
-.1
-.2
-.3
-.4
-.5
-.6

Figure 1

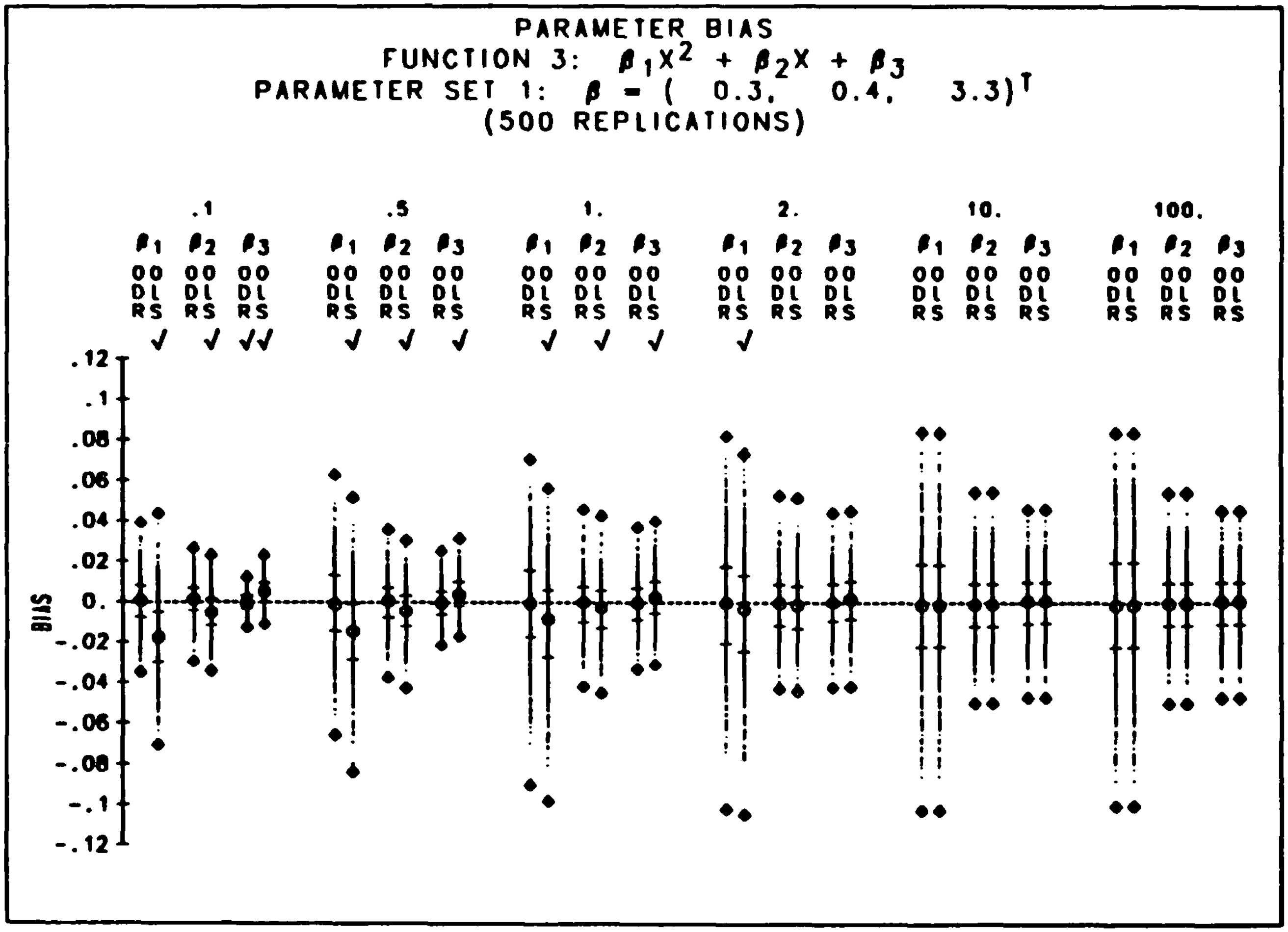

PARAMETER BIAS
FUNCTION 3: $\beta_1X^2 + \beta_2X + \beta_3$
PARAMETER SET 1: $\beta = (0.3, 0.4, 3.3)^T$
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
β_1
β_2
β_3
OO DL RS
✓
✓✓
BIAS
.12
.1
.08
.06
.04
.02
0.
-.02
-.04
-.06
-.08
-.1
-.12

Figure 2

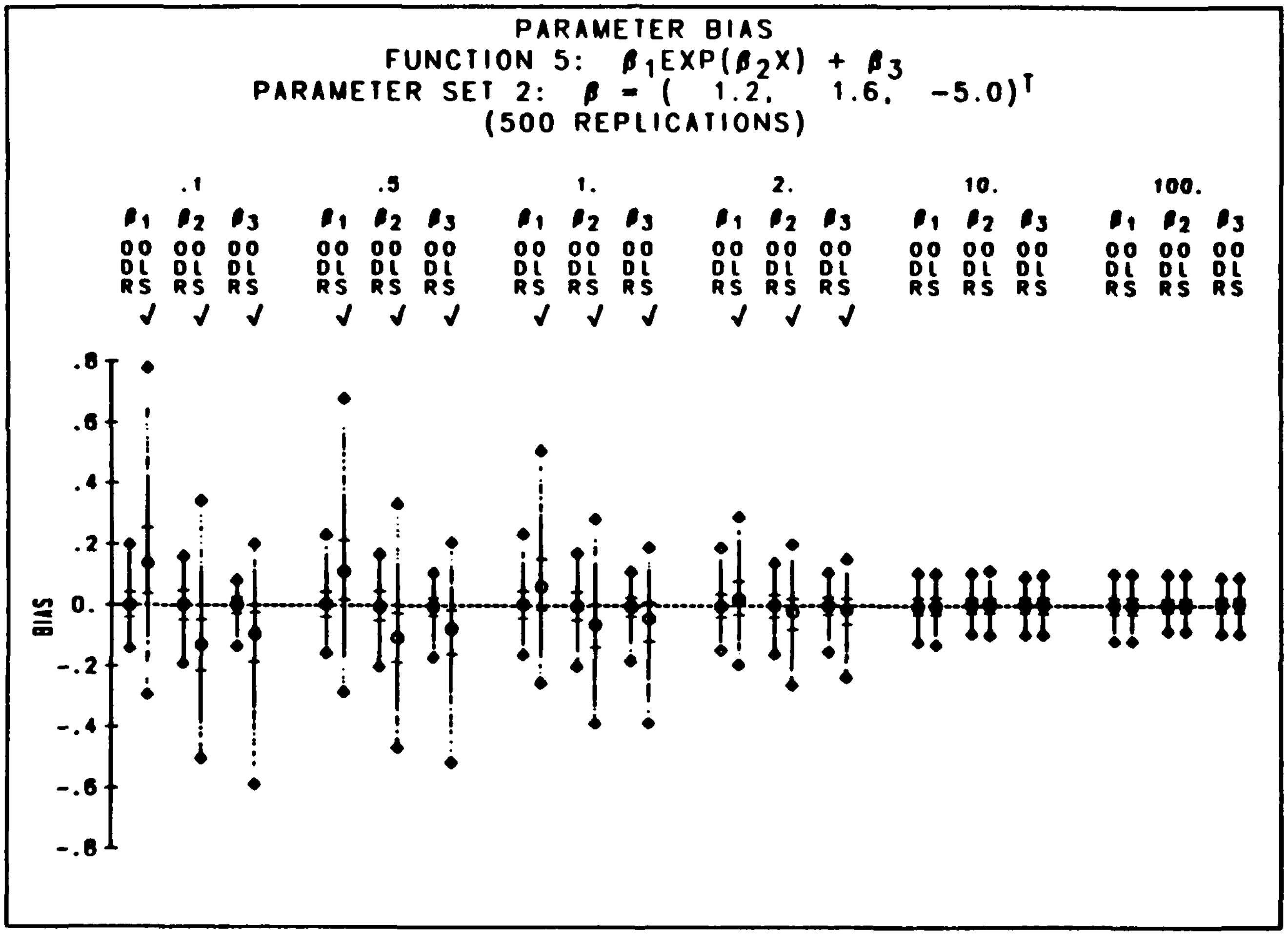

PARAMETER BIAS
FUNCTION 5: $\beta_1 EXP(\beta_2 X) + \beta_3$
PARAMETER SET 2: $\beta = (1.2, 1.6, -5.0)^T$
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
β_1
β_2
β_3
OO
DL
RS
√
BIAS
.8
.6
.4
.2
0.
-.2
-.4
-.6
-.8

Figure 3

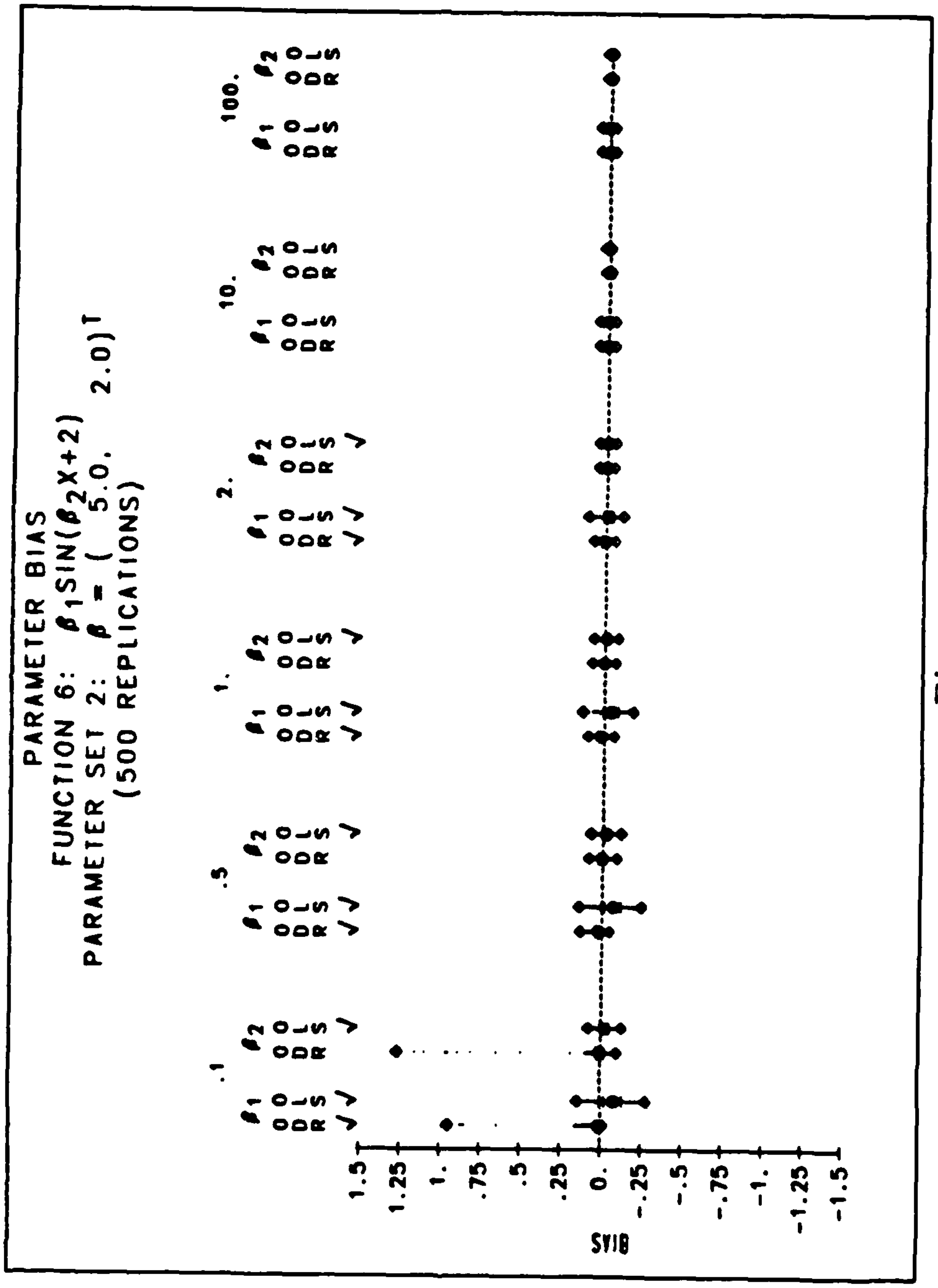
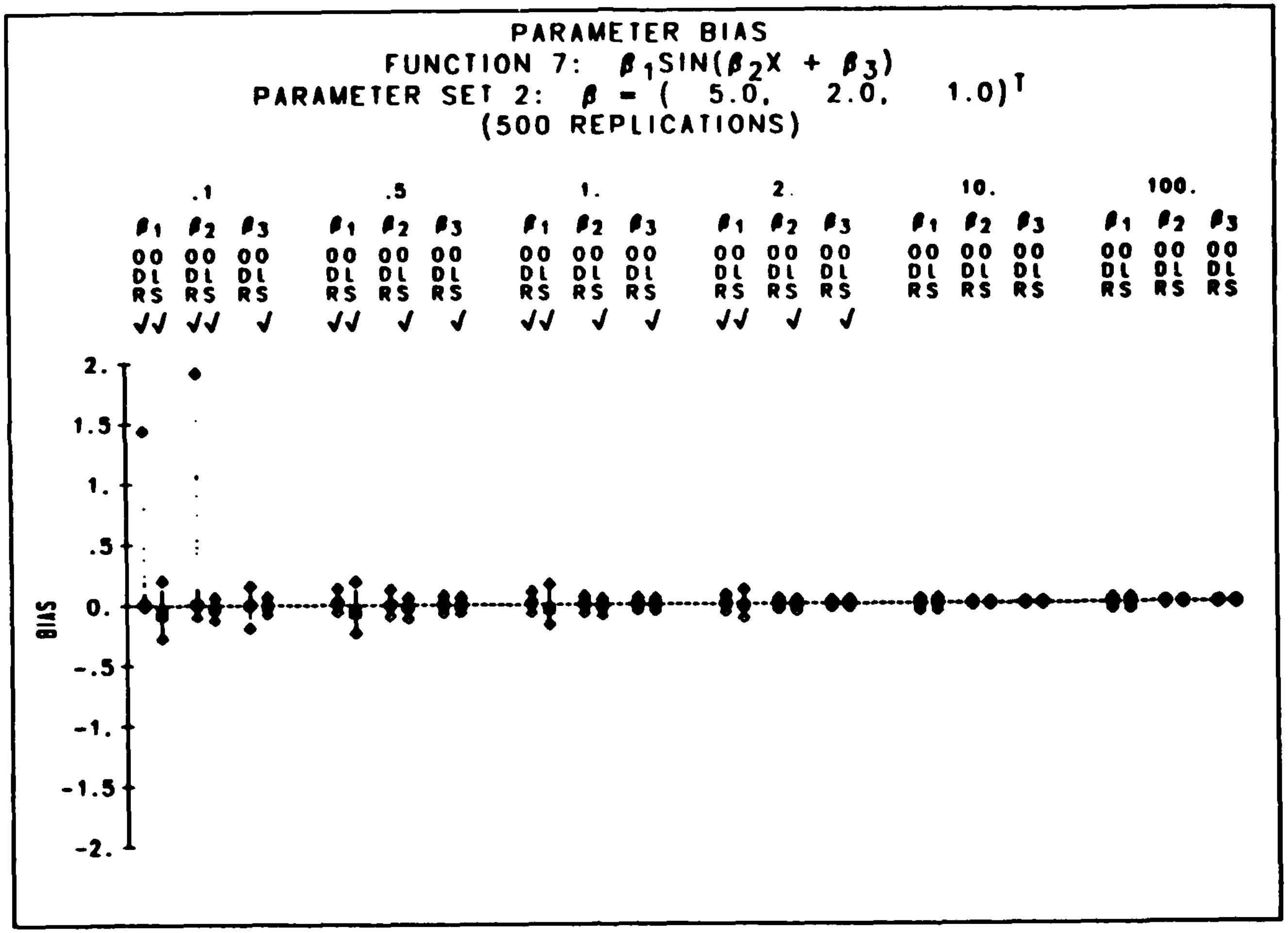

PARAMETER BIAS
FUNCTION 6: $\beta_1 SIN(\beta_2 X+2)$
PARAMETER SET 2: β = (5.0, 2.0)^T
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
β_1
β_2
OLS
ODR
BIAS
1.5
1.25
1.
.75
.5
.25
0.
-.25
-.5
-.75
-1.
-1.25
-1.5

Figure 4



PARAMETER BIAS
FUNCTION 7: $\beta_1 \mathrm{SIN}(\beta_2 X + \beta_3)$
PARAMETER SET 2: $\beta = (5.0, \; 2.0, \; 1.0)^T$
(500 REPLICATIONS)

Figure 5

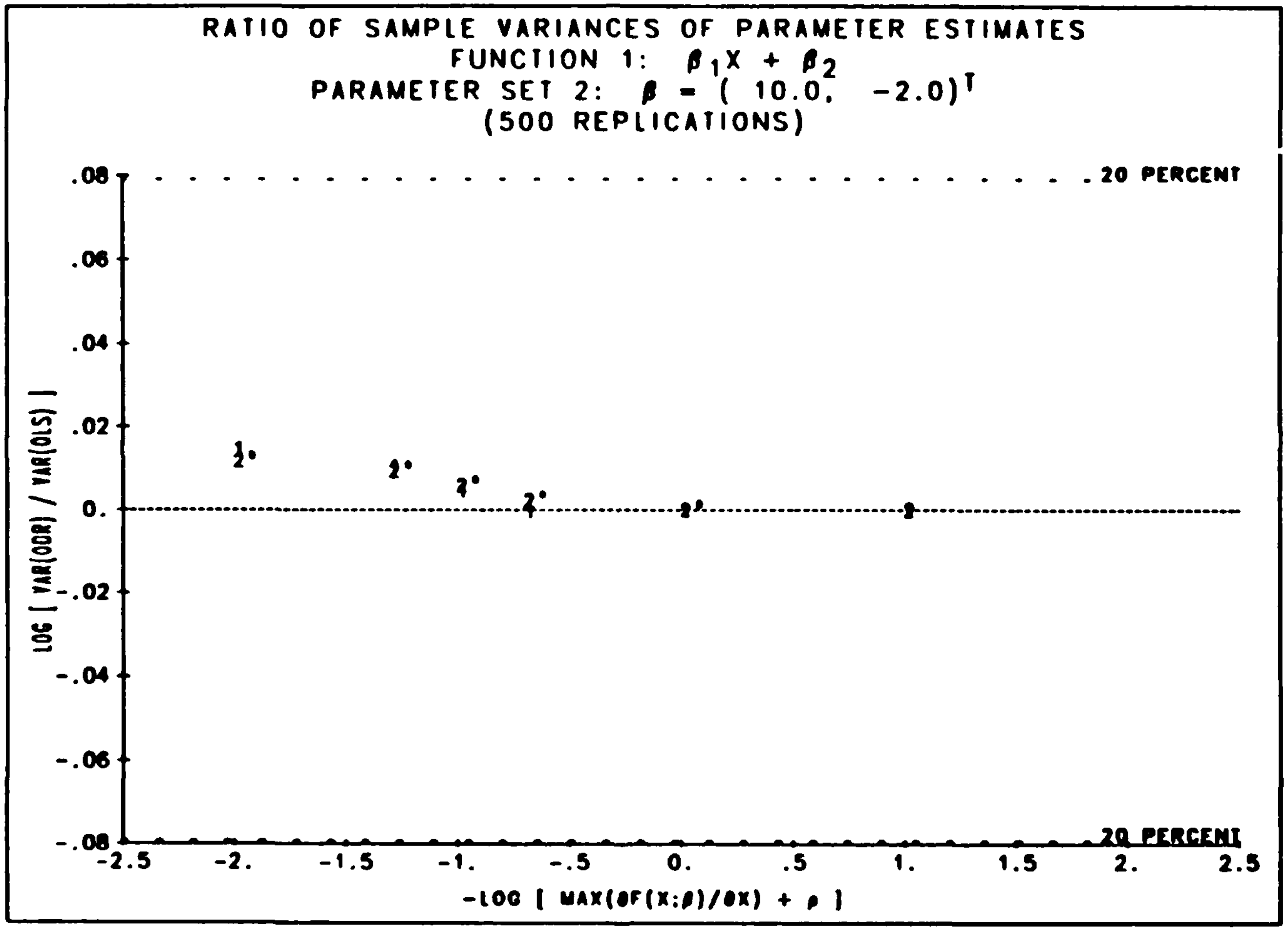

RATIO OF SAMPLE VARIANCES OF PARAMETER ESTIMATES
FUNCTION 1: $\beta_1 X + \beta_2$
PARAMETER SET 2: $\beta = (\ 10.0,\ \ -2.0)^T$
(500 REPLICATIONS)
.08
.06
.04
.02
0.
-.02
-.04
-.06
-.08
20 PERCENT
20 PERCENT
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
LOG [VAR(ODR) / VAR(OLS)]
-LOG [MAX(∂F(X;β)/∂X) + ρ]
1
2
2
2
2
2
2

Figure 6

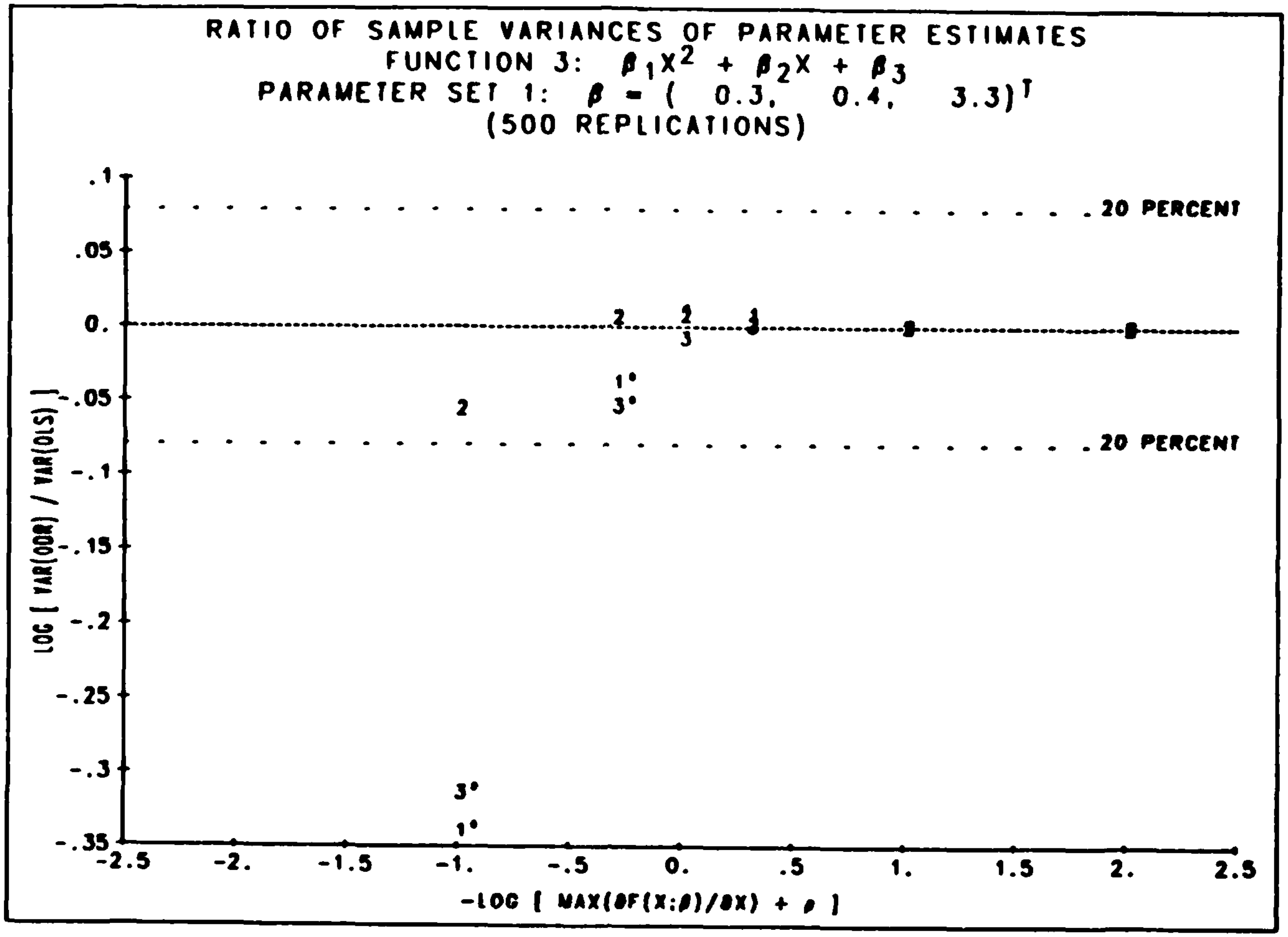

RATIO OF SAMPLE VARIANCES OF PARAMETER ESTIMATES
FUNCTION 3: $\beta_1 X^2 + \beta_2 X + \beta_3$
PARAMETER SET 1: $\beta = (\ 0.3,\ \ 0.4,\ \ 3.3)^T$
(500 REPLICATIONS)
.1
.05
0.
-.05
-.1
-.15
-.2
-.25
-.3
-.35
LOG [VAR(ODR) / VAR(OLS)]
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X;β)/∂X) + ρ]
20 PERCENT
20 PERCENT
2
1*
3*
3*
1*
1
2
3

Figure 7

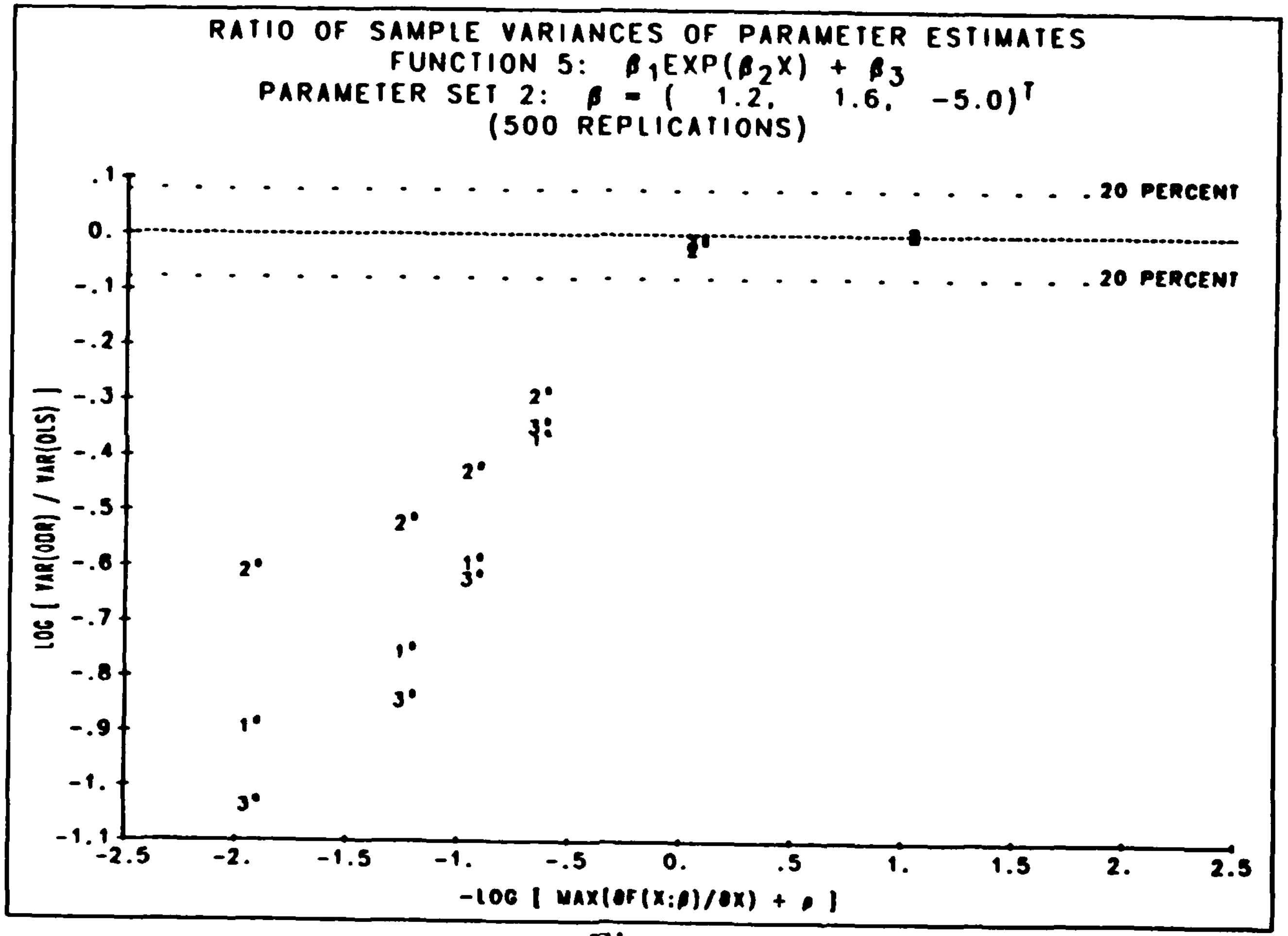

RATIO OF SAMPLE VARIANCES OF PARAMETER ESTIMATES
FUNCTION 5: $\beta_1 EXP(\beta_2 X) + \beta_3$
PARAMETER SET 2: $\beta = (1.2, 1.6, -5.0)^T$
(500 REPLICATIONS)
.1
0.
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.
-1.1
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
.20 PERCENT
.20 PERCENT
LOG [VAR(ODR) / VAR(OLS)]
-LOG [MAX(∂F(X;β)/∂X) + ρ]
1
2
3

Figure 8

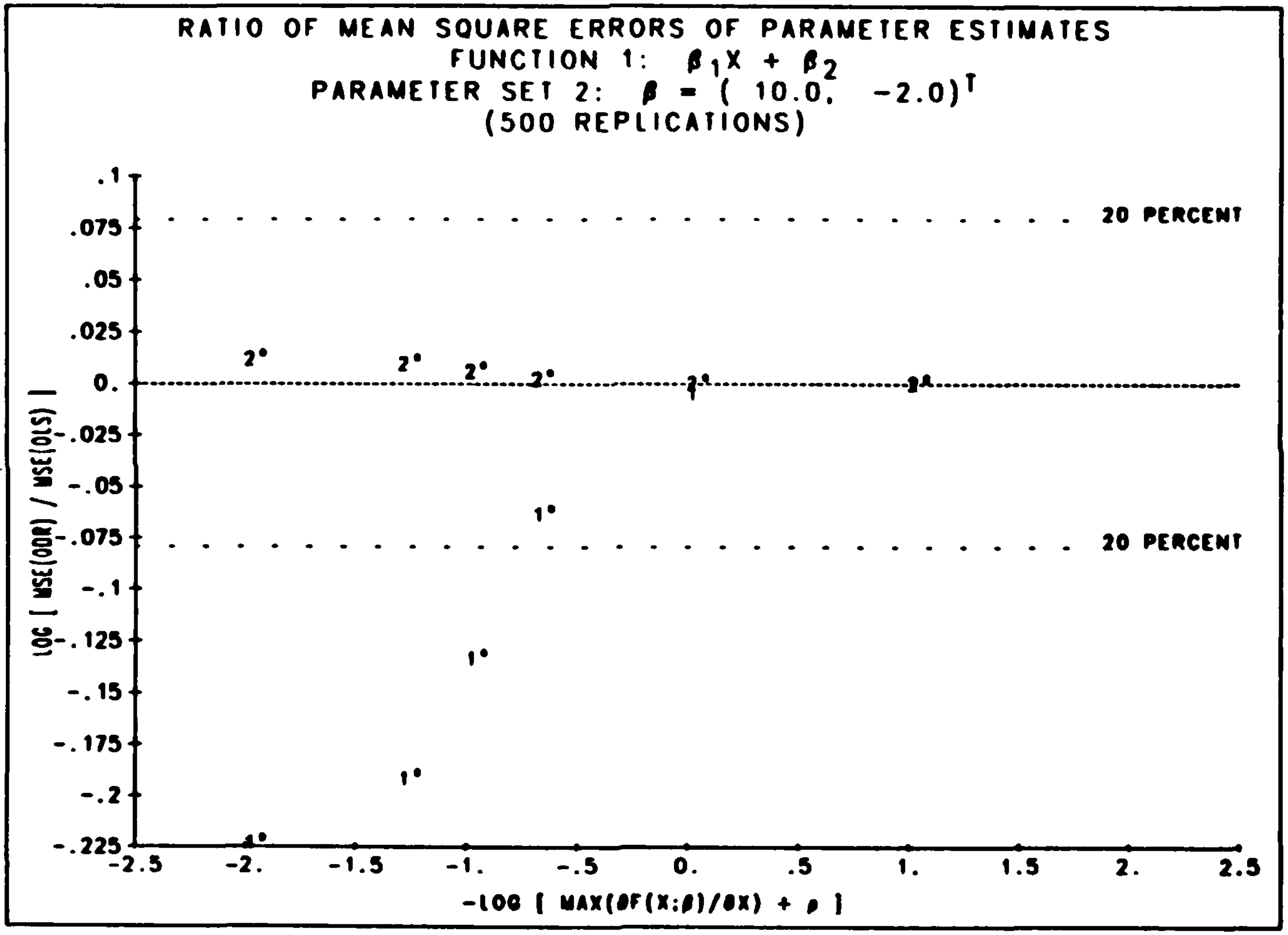

RATIO OF MEAN SQUARE ERRORS OF PARAMETER ESTIMATES
FUNCTION 1: $\beta_1 X + \beta_2$
PARAMETER SET 2: $\beta = (10.0, -2.0)^T$
(500 REPLICATIONS)
LOG [MSE(ODR) / MSE(OLS)]
.1
.075
.05
.025
0.
-.025
-.05
-.075
-.1
-.125
-.15
-.175
-.2
-.225
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X:β)/∂X) + β]
20 PERCENT
20 PERCENT
1
2

Figure 9

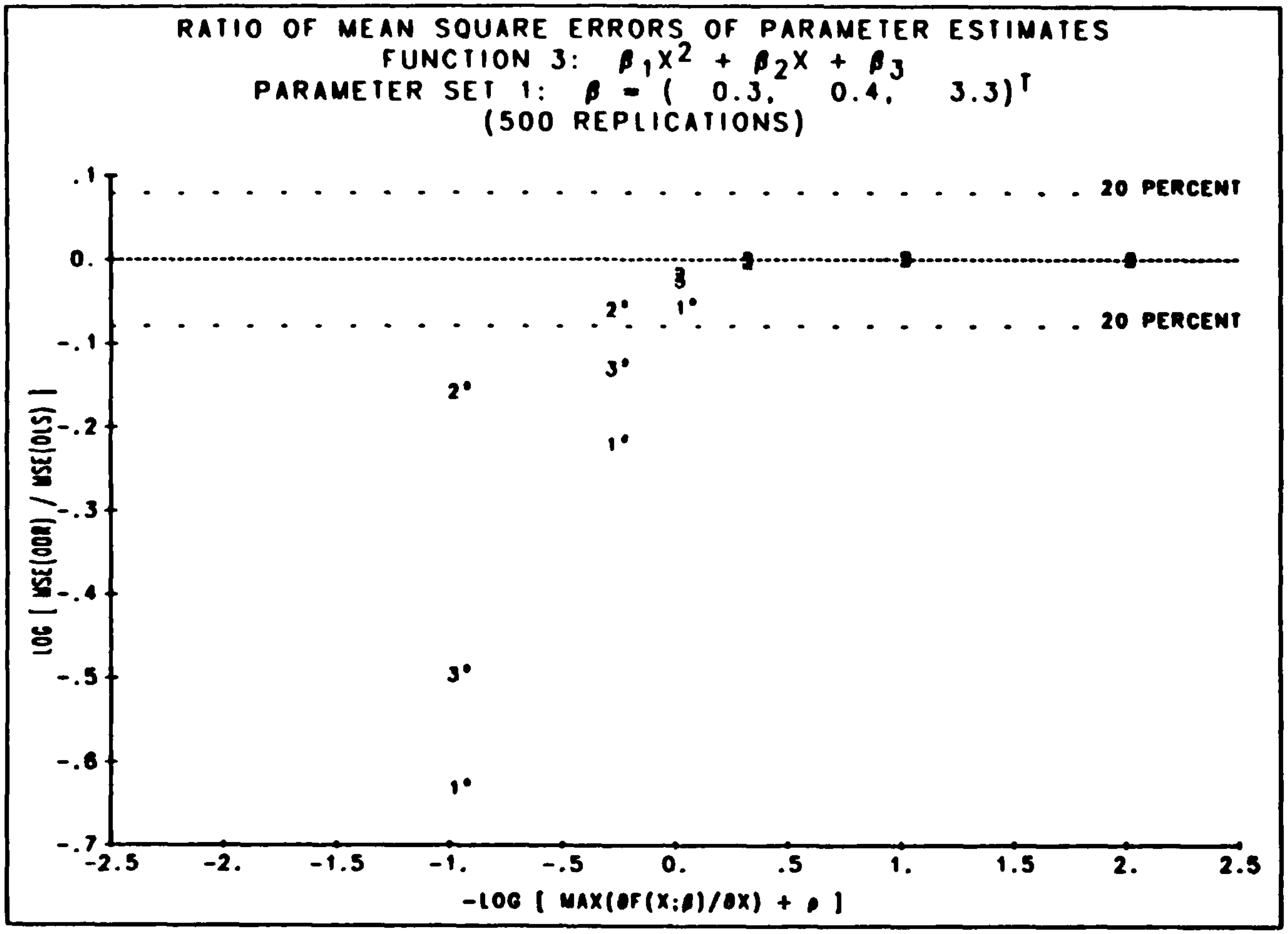

RATIO OF MEAN SQUARE ERRORS OF PARAMETER ESTIMATES
FUNCTION 3: $\beta_1 X^2 + \beta_2 X + \beta_3$
PARAMETER SET 1: $\beta = (\ 0.3,\ \ 0.4,\ \ 3.3)^T$
(500 REPLICATIONS)
20 PERCENT
20 PERCENT
.1
0.
-.1
-.2
-.3
-.4
-.5
-.6
-.7
LOG [MSE(ODR) / MSE(OLS)]
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X;β)/∂X) + ρ]
3
2*
1*
3*
1*
2*
3*
1*

Figure 10

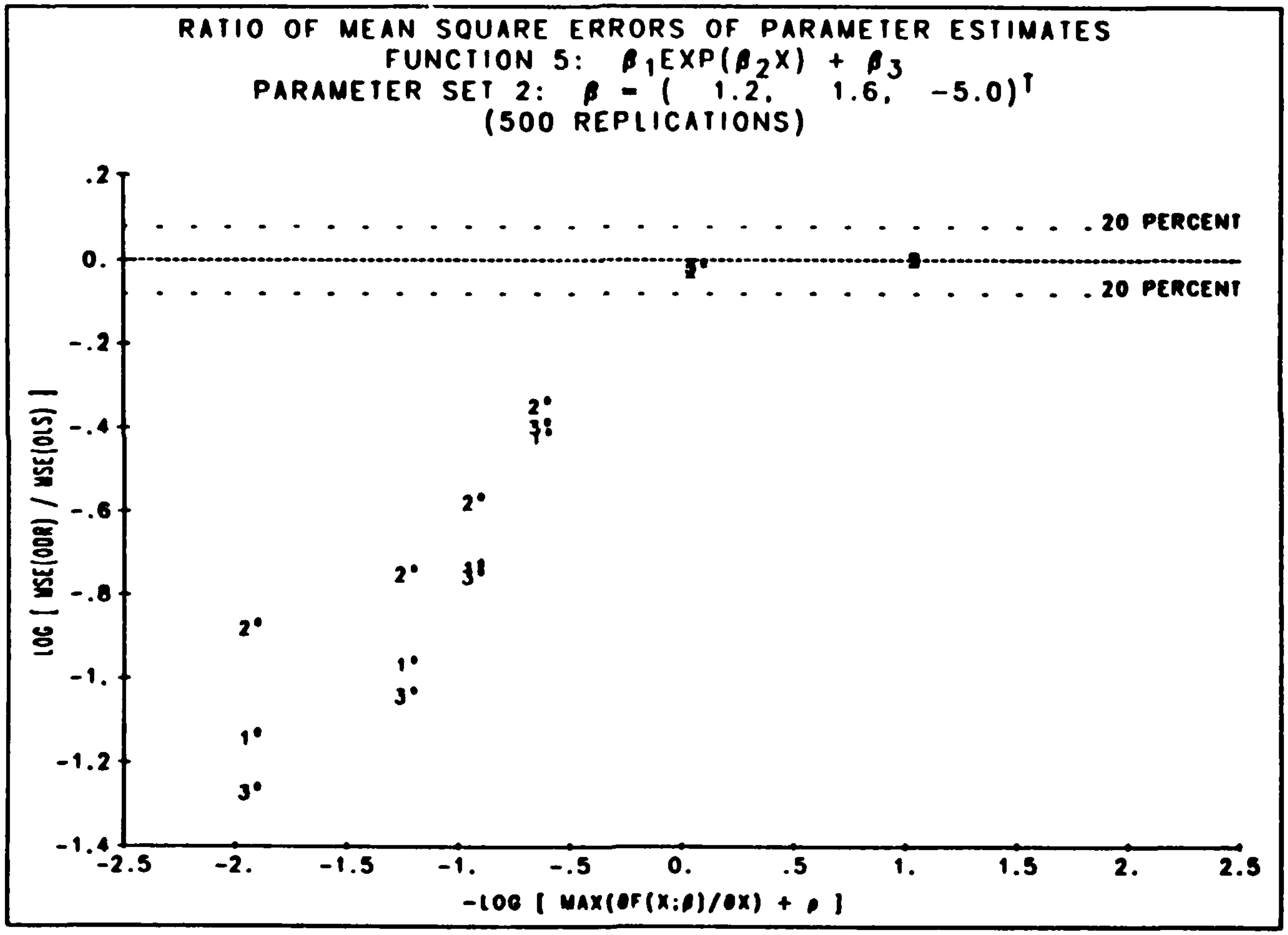

RATIO OF MEAN SQUARE ERRORS OF PARAMETER ESTIMATES
FUNCTION 5: $\beta_1 EXP(\beta_2 X) + \beta_3$
PARAMETER SET 2: $\beta = (1.2, 1.6, -5.0)^T$
(500 REPLICATIONS)
.2
0.
-.2
-.4
-.6
-.8
-1.
-1.2
-1.4
LOG [MSE(ODR) / MSE(OLS)]
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X;β)/∂X) + ρ]
20 PERCENT
20 PERCENT
1
2
3

Figure 11

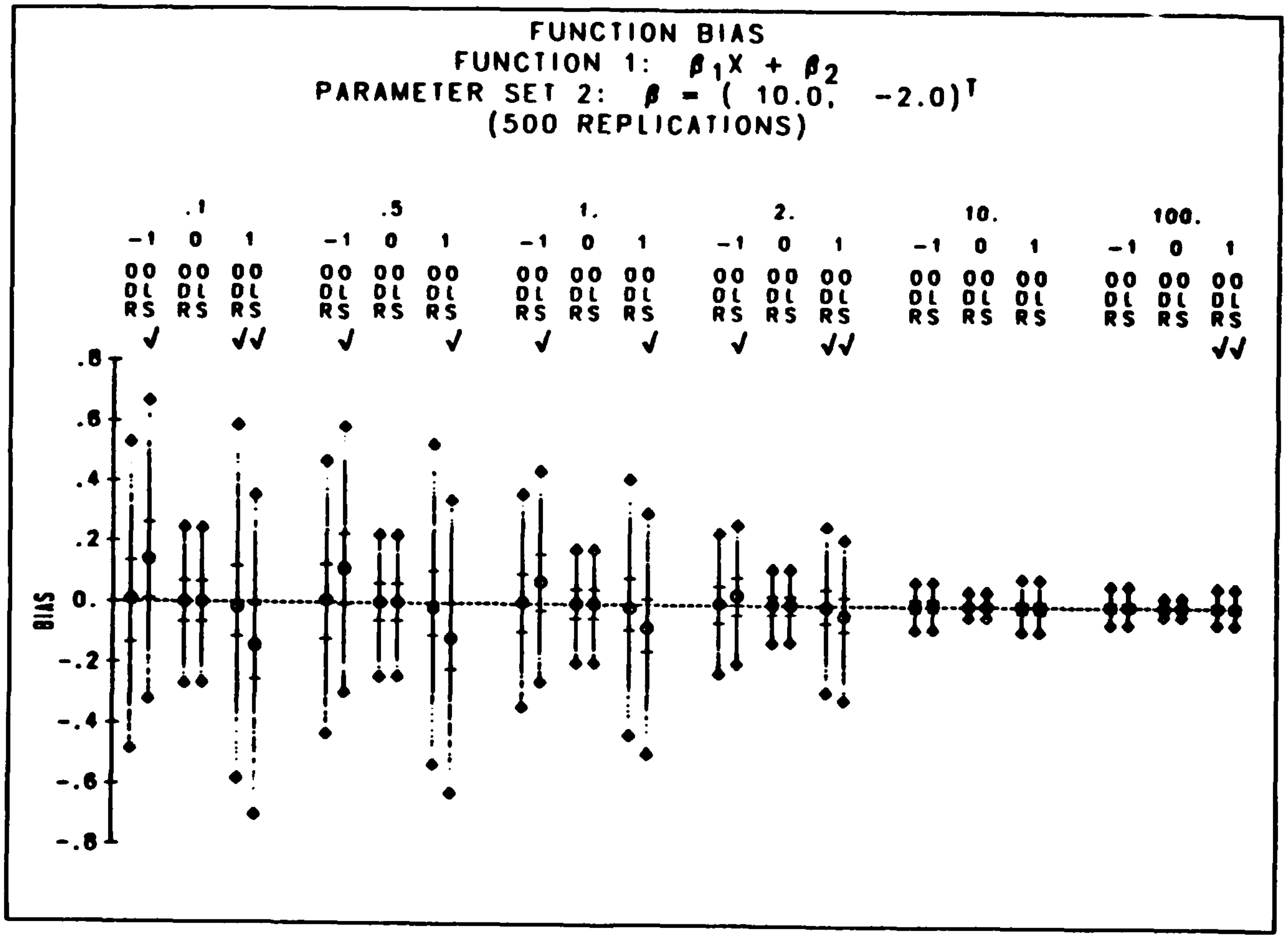

FUNCTION BIAS
FUNCTION 1: $\beta_1 X + \beta_2$
PARAMETER SET 2: $\beta = (10.0, -2.0)^T$
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
-1 0 1
OO
DL
RS
√
√√
BIAS
.8
.6
.4
.2
0.
-.2
-.4
-.6
-.8

Figure 12

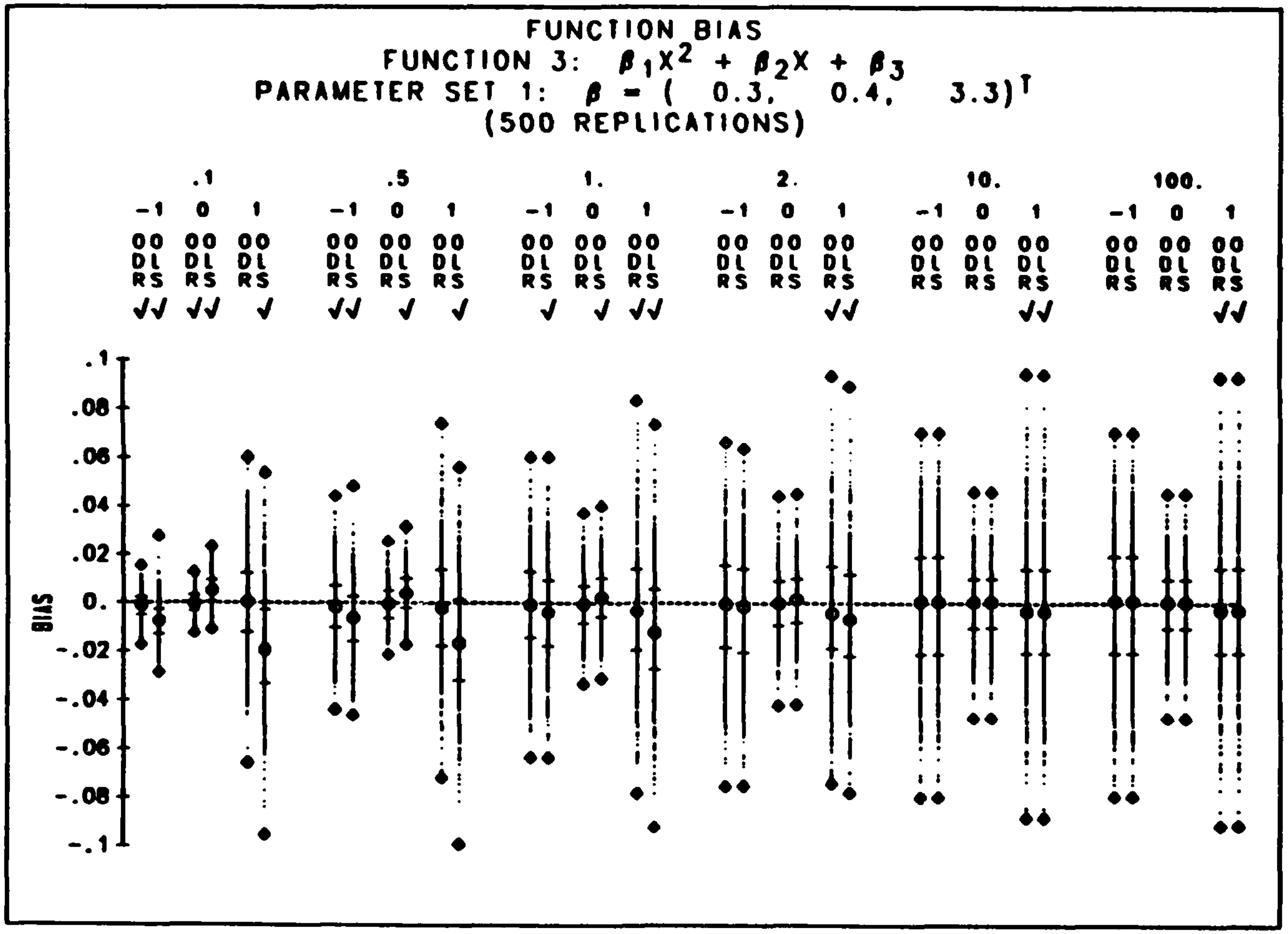

FUNCTION BIAS
FUNCTION 3: $\beta_1 X^2 + \beta_2 X + \beta_3$
PARAMETER SET 1: β = (0.3, 0.4, 3.3)^T
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
-1
0
1
OO
DL
RS
√√
√
BIAS
.1
.08
.06
.04
.02
0.
-.02
-.04
-.06
-.08
-.1

Figure 13

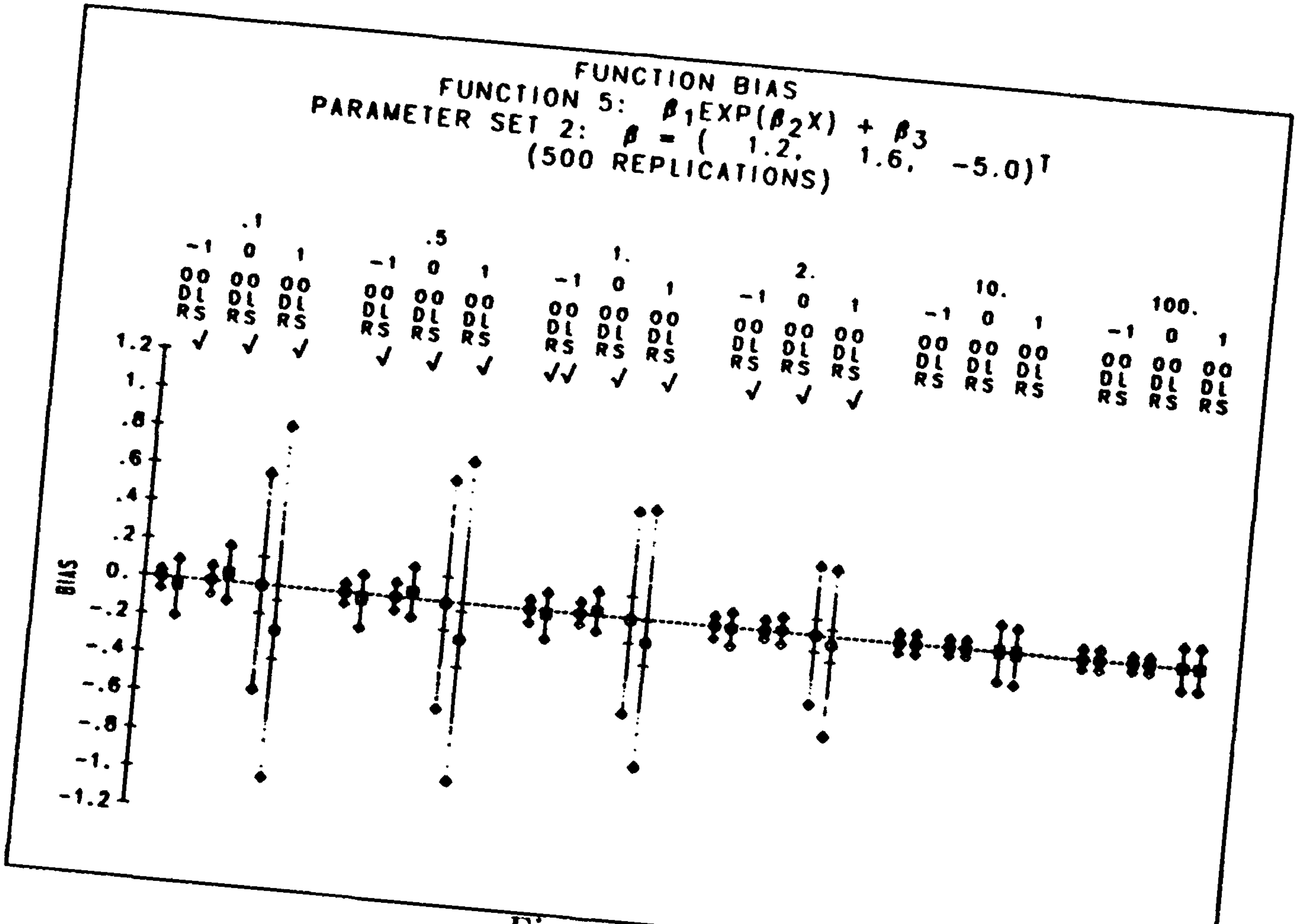

FUNCTION BIAS
FUNCTION 5: β1EXP(β2X) + β3
PARAMETER SET 2: β = (1.2, 1.6, -5.0)T
(500 REPLICATIONS)
.1
.5
1.
2.
10.
100.
-1
0
1
OO
DL
RS
√
√√
BIAS
1.2
1.
.8
.6
.4
.2
0.
-.2
-.4
-.6
-.8
-1.
-1.2

Figure 14

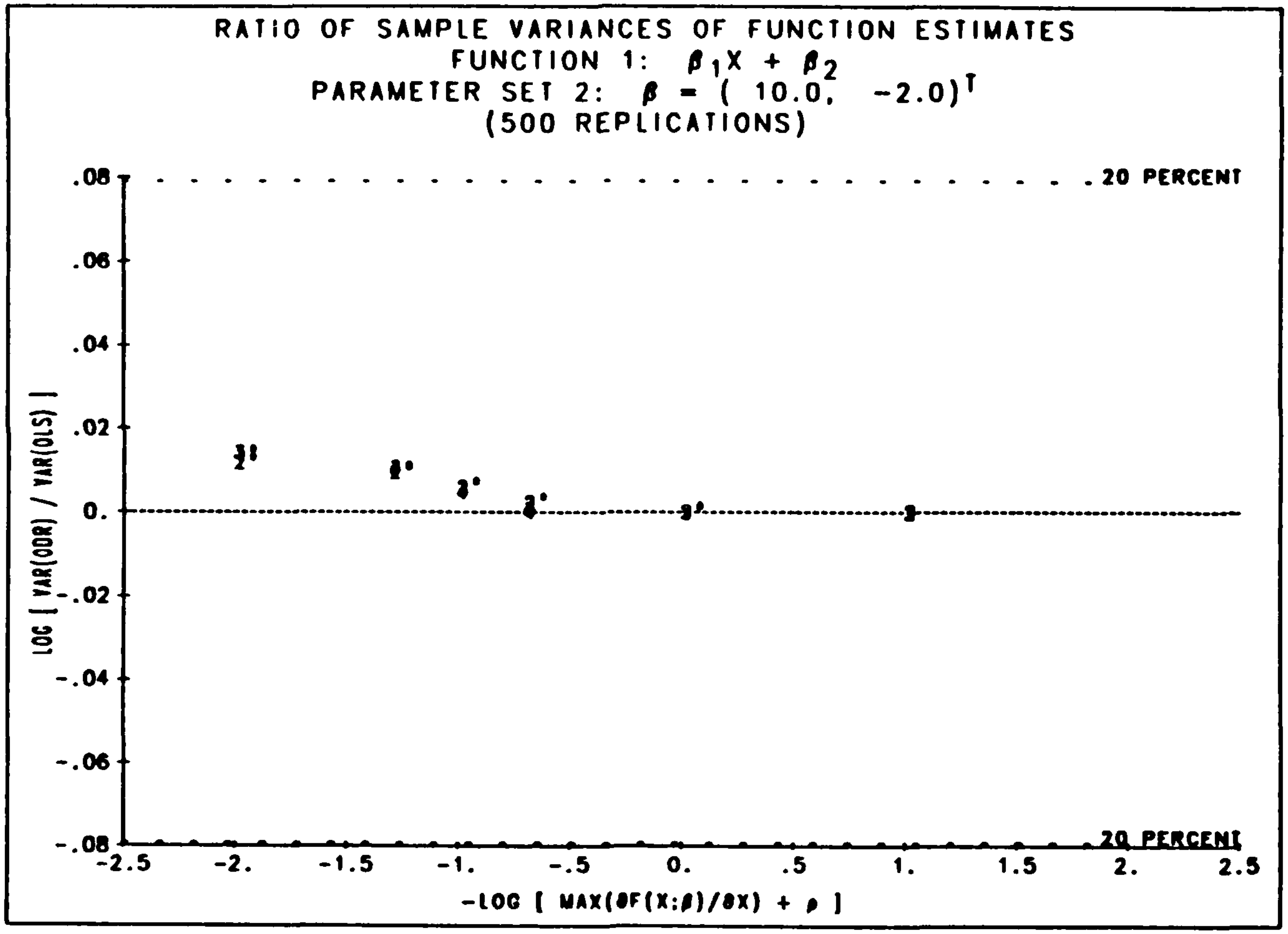

RATIO OF SAMPLE VARIANCES OF FUNCTION ESTIMATES
FUNCTION 1: $\beta_1 X + \beta_2$
PARAMETER SET 2: $\beta = (\ 10.0,\ \ -2.0)^T$
(500 REPLICATIONS)
.08
.06
.04
.02
0.
-.02
-.04
-.06
-.08
20 PERCENT
20 PERCENT
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
LOG [VAR(ODR) / VAR(OLS)]
-LOG [MAX(∂F(X;β)/∂X) + ρ]

Figure 15

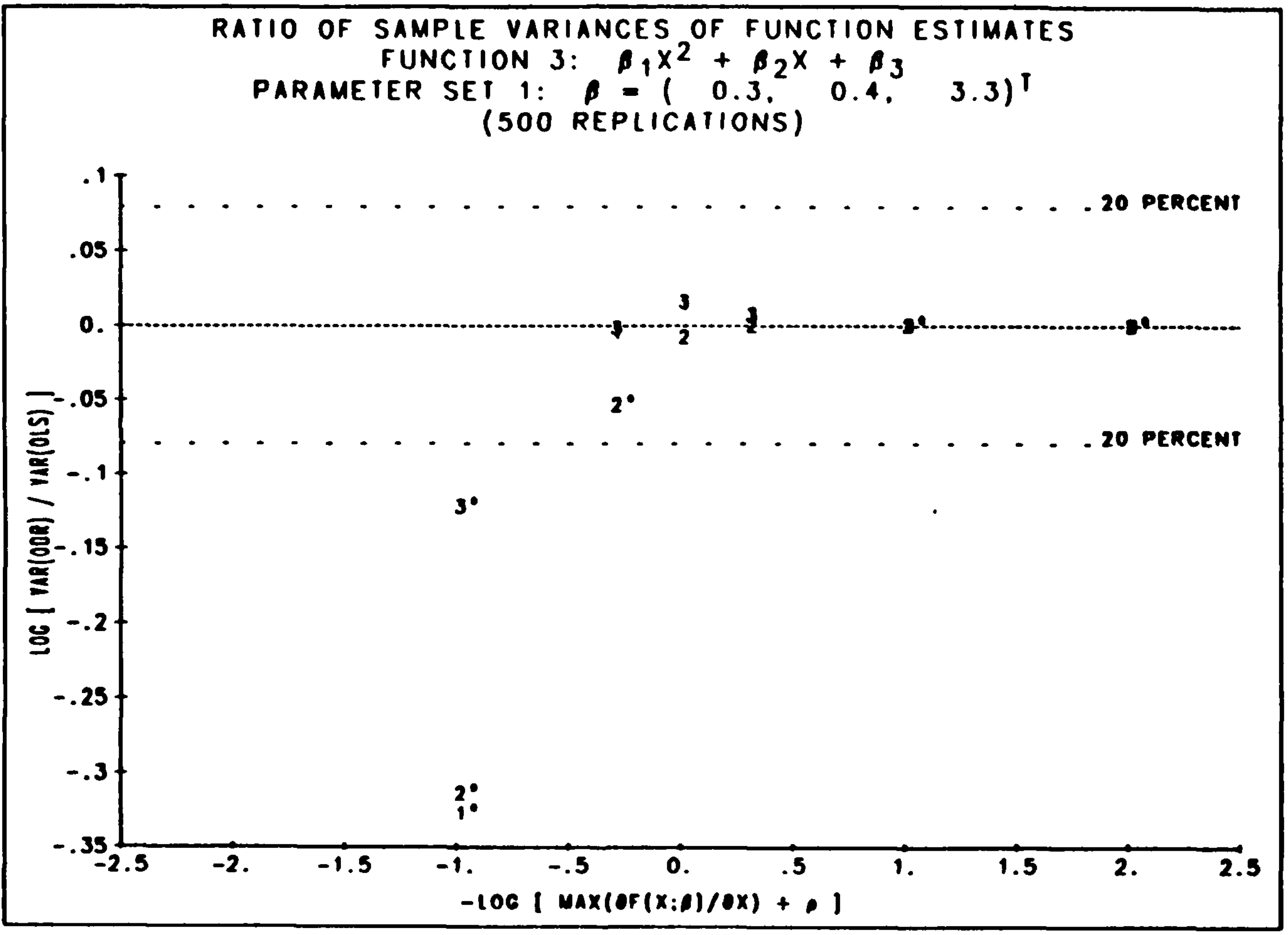

RATIO OF SAMPLE VARIANCES OF FUNCTION ESTIMATES
FUNCTION 3: $\beta_1 X^2 + \beta_2 X + \beta_3$
PARAMETER SET 1: $\beta = (0.3, 0.4, 3.3)^T$
(500 REPLICATIONS)
.1
.05
0.
-.05
-.1
-.15
-.2
-.25
-.3
-.35
LOG [VAR(ODR) / VAR(OLS)]
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X:β)/∂X) + ρ]
20 PERCENT
20 PERCENT
3
2
2*
3*
2*
1*

Figure 16

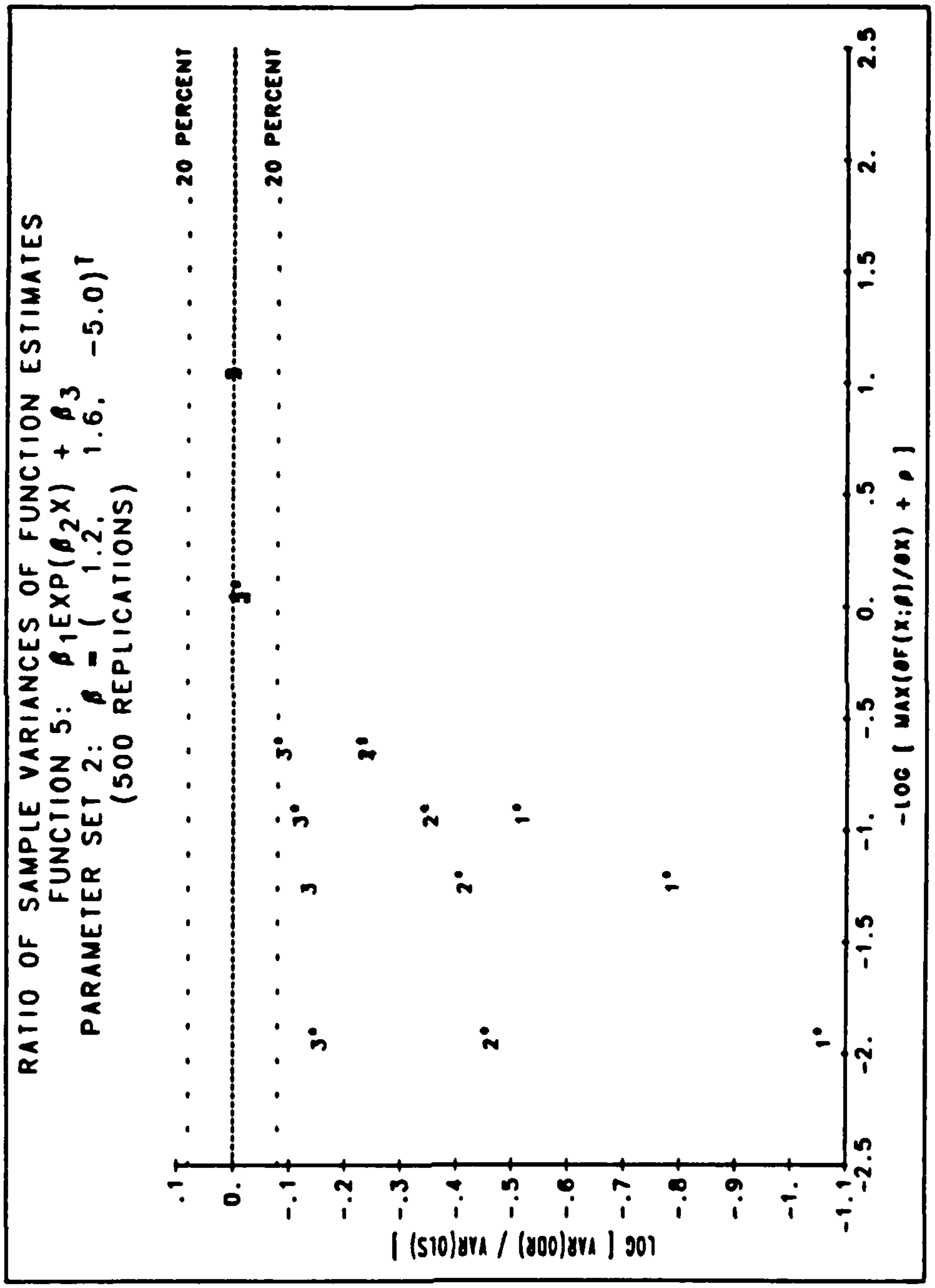

RATIO OF SAMPLE VARIANCES OF FUNCTION ESTIMATES
FUNCTION 5: $\beta_1 EXP(\beta_2 X) + \beta_3$
PARAMETER SET 2: $\beta = (1.2, 1.6, -5.0)^T$
(500 REPLICATIONS)
.20 PERCENT
.20 PERCENT
LOG [VAR(ODR) / VAR(OLS)]
.1
0.
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-.8
-.9
-1.
-1.1
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X;β)/∂X) + ρ]
1*
2*
3*
1*
2*
3
1*
2*
3*
2*
3*
1*
1*

Figure 17

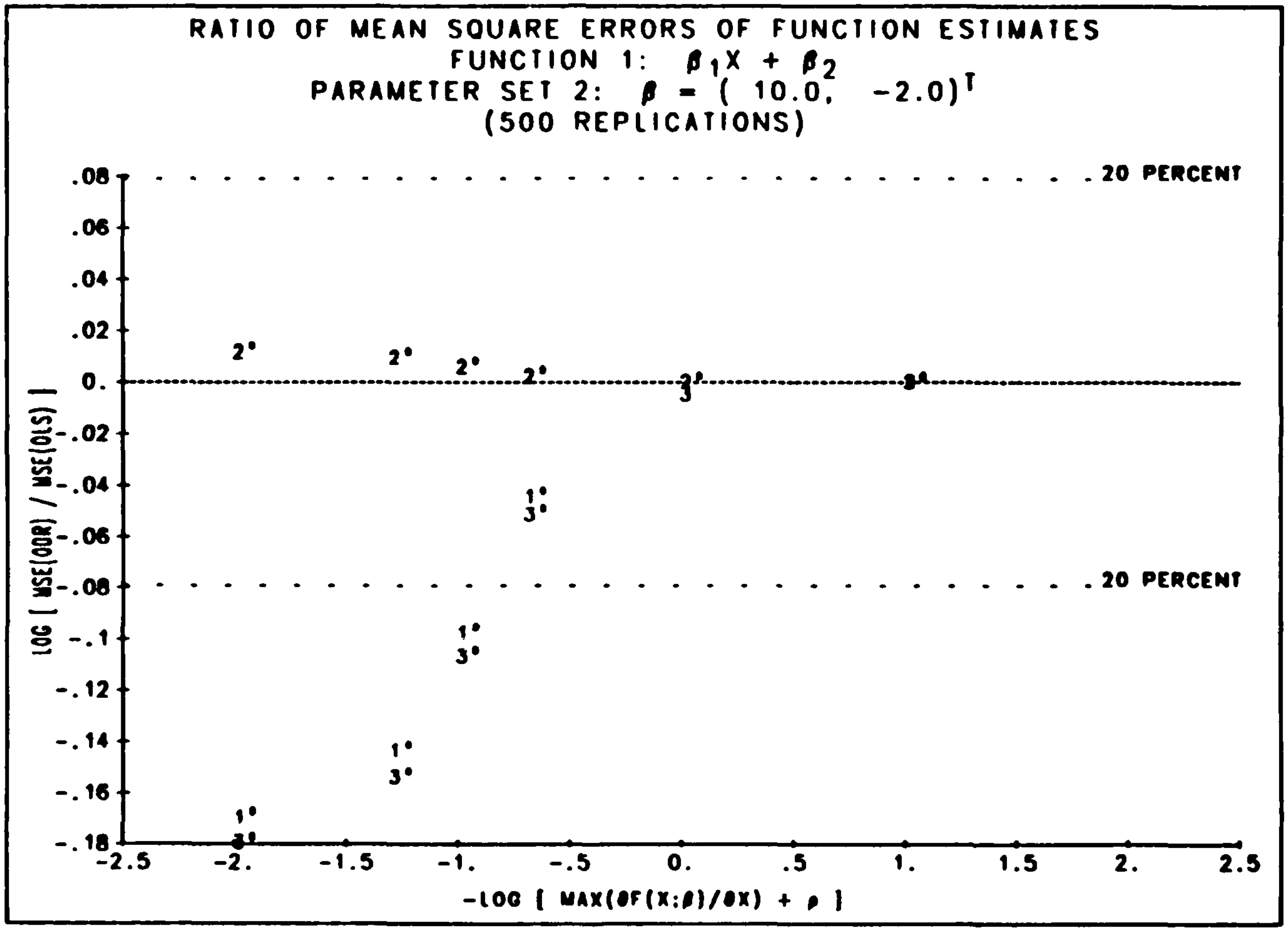

RATIO OF MEAN SQUARE ERRORS OF FUNCTION ESTIMATES
FUNCTION 1: $\beta_1 X + \beta_2$
PARAMETER SET 2: $\beta = (10.0, -2.0)^T$
(500 REPLICATIONS)
.08
.06
.04
.02
0.
-.02
-.04
-.06
-.08
-.1
-.12
-.14
-.16
-.18
20 PERCENT
20 PERCENT
1*
2*
3*
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
LOG [MSE(ODR) / MSE(OLS)]
-LOG [MAX(∂F(X;β)/∂X) + ρ]

Figure 18

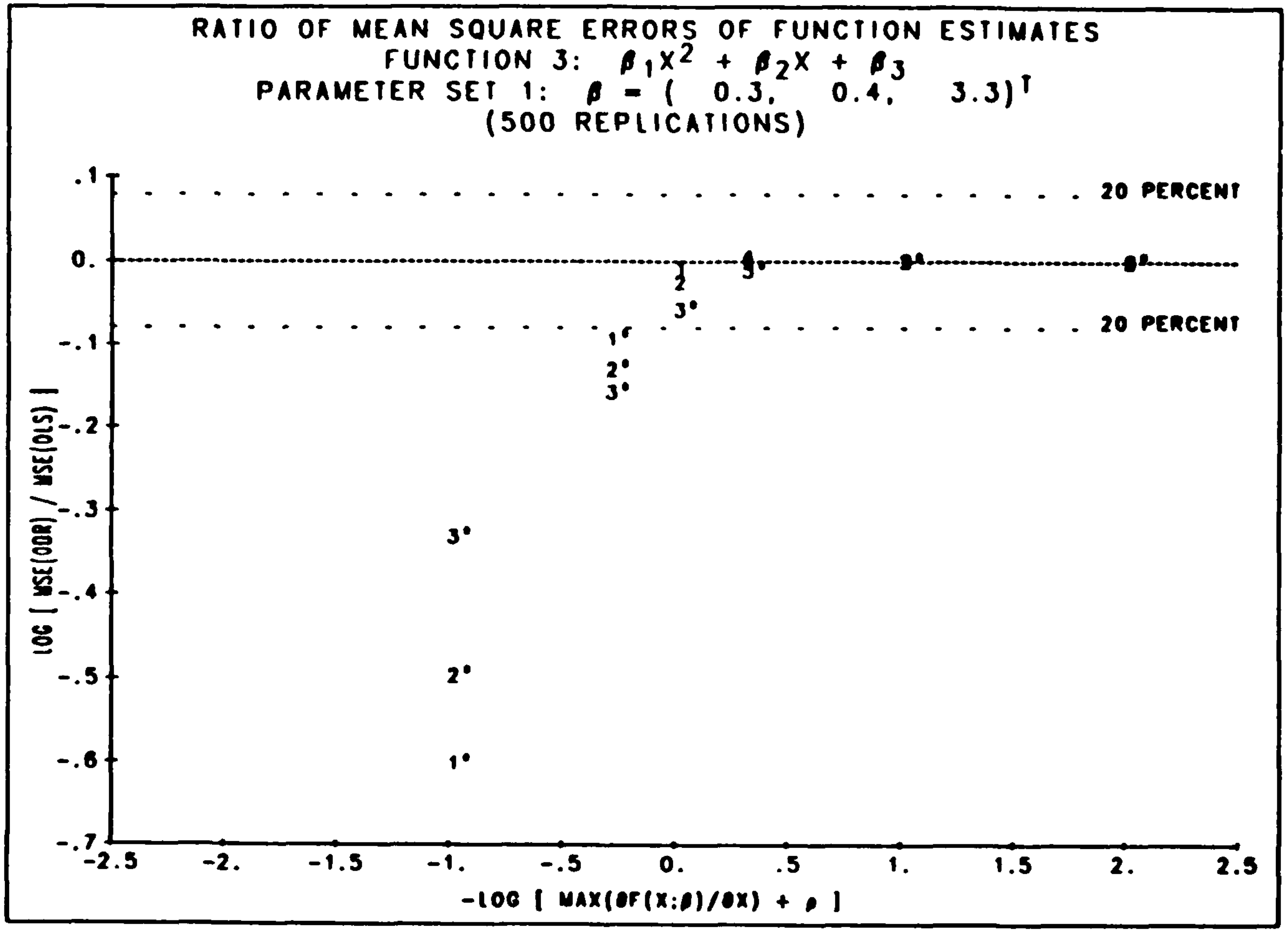

RATIO OF MEAN SQUARE ERRORS OF FUNCTION ESTIMATES
FUNCTION 3: $\beta_1 X^2 + \beta_2 X + \beta_3$
PARAMETER SET 1: $\beta = (0.3, 0.4, 3.3)^T$
(500 REPLICATIONS)
20 PERCENT
20 PERCENT
.1
0.
-.1
-.2
-.3
-.4
-.5
-.6
-.7
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
LOG [MSE(ODR) / MSE(OLS)]
-LOG [MAX(∂F(X;β)/∂X) + ρ]
1
2
3

Figure 19

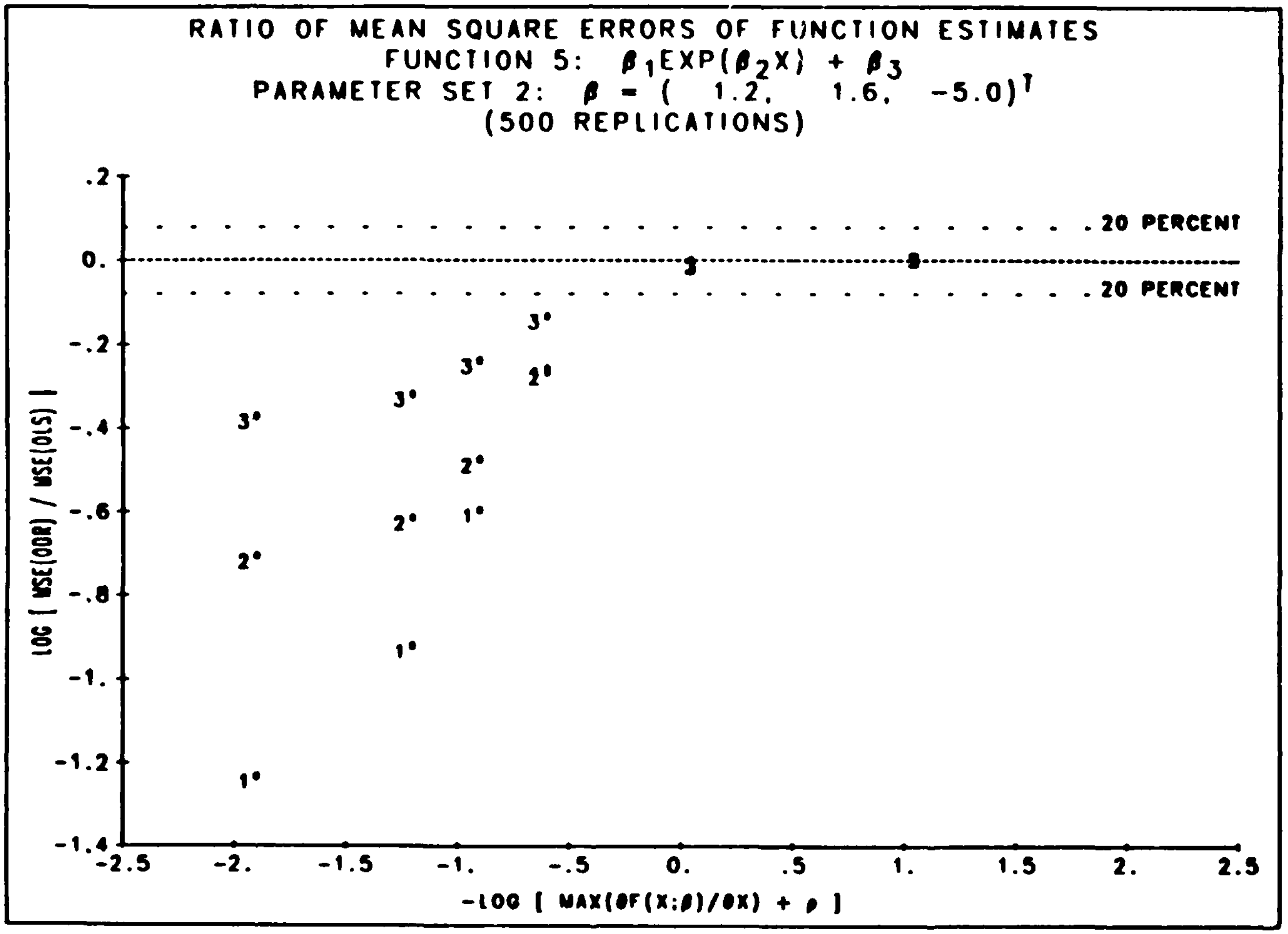

RATIO OF MEAN SQUARE ERRORS OF FUNCTION ESTIMATES
FUNCTION 5: $\beta_1 EXP(\beta_2 X) + \beta_3$
PARAMETER SET 2: $\beta = (1.2, 1.6, -5.0)^T$
(500 REPLICATIONS)
.2
0.
-.2
-.4
-.6
-.8
-1.
-1.2
-1.4
LOG [MSE(ODR) / MSE(OLS)]
20 PERCENT
20 PERCENT
1*
2*
3*
-2.5
-2.
-1.5
-1.
-.5
0.
.5
1.
1.5
2.
2.5
-LOG [MAX(∂F(X;β)/∂X) + ρ]

Figure 20

References

[BogDSS87] Boggs, P. T., J. R. Donaldson, R. B. Schnabel and C. H. Spiegelman (1987), "Orthogonal Distance Regression vs Ordinary Least Squares — A Monte Carlo Study." To appear as an National Bureau of Standards Internal Report.

[BogBDS87] Boggs, P. T., R. H. Byrd, J. R. Donaldson and R. B. Schnabel (1987), "ODRPACK — Software for Weighted Orthogonal Distance Regression." University of Colorado Department of Computer Science Technical Report Number CU-CS-360-87.

[BogBS85] Boggs, P. T., R. H. Byrd, and R. B. Schnabel (1985), "A Stable and Efficient Algorithm for Nonlinear Orthogonal Distance Regression." University of Colorado Department of Computer Science Technical Report Number CU-CS-317-85. (To appear in *SIAM J. Sci. Stat. Computing*, 1987.)

[GolV83] Golub, G. H. and C. F. VanLoan (1983), *Matrix Computations*, Johns Hopkins University Press, Baltimore, Maryland.

[Ful87] Fuller, W. A. (1987), *Measurement Error Models*, John Wiley and Sons, New York, New York.

[MarT84] Marsaglia, G., and W. W. Tsang (1984), "A Fast, Easily Implemented Method for Sampling From Decreasing or Symmetric Unimodal Density Functions," *SIAM J. Sci. Stat. Computing*, 5(2): 349-359.

[Meg86] Megatex (1986). *Template V5.5 2D/3D Reference Manual*, Megatex Corporation, San Diego, California.

[Mor71] Morran, P. (1971). "Estimating Structural and Functional Relationships." *Journal of Multivariate Analysis*, 1(2): 232-255.

[Rao86] Rao, C. R. (1986), "The Pitman Nearness Criteria and its Determination," *Communications in Statistics: Theory and Methods*, 15(11): 3173-3191.

[ReiGL86] Reilman, M. A., R. F. Gunst and M. Y. Lakshminarayanan (1986). "Stochastic Regression with Errors in Both Variables," *Journal of Quality Technology*. 18(3): 162-169.

[Tuk77] Tukey. J. W. (1977), *Exploratory Data Analysis*, Addison-Wesley. Reading, Massachusetts.

ADA184374

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION
Unclassified

1b. RESTRICTIVE MARKINGS

2a. SECURITY CLASSIFICATION AUTHORITY

2b. DECLASSIFICATION/DOWNGRADING SCHEDULE

3. DISTRIBUTION/AVAILABILITY OF REPORT
Approved for public release;
Distribution unlimited

4. PERFORMING ORGANIZATION REPORT NUMBER(S)
CU-CS-362-87

5. MONITORING ORGANIZATION REPORT NUMBER(S)
ARO 21453.10-MA

6a. NAME OF PERFORMING ORGANIZATION
University of Colorado

6b. OFFICE SYMBOL (If applicable)

7a. NAME OF MONITORING ORGANIZATION
U.S. Army Research Office

6c. ADDRESS (City, State and ZIP Code)
Computer Science Department
Campus Box 430
Boulder, CO 80309-0430

7b. ADDRESS (City, State and ZIP Code)
Post Office Box 12211
Rèsearch Triangle Park, NC 27709

8a. NAME OF FUNDING/SPONSORING ORGANIZATION

8b. OFFICE SYMBOL (If applicable)

9. PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER
DAAG-29-84-K-0140

8c. ADDRESS (City, State and ZIP Code)

10. SOURCE OF FUNDING NOS.

PROGRAM ELEMENT NO.	PROJECT NO.	TASK NO.	WORK UNIT NO.

11. TITLE (Include Security Classification)
A Computational Examination of Orthogonal Distance Regression

12. PERSONAL AUTHOR(S)
Paul T. Boggs, Janet R. Donaldson, Robert B. Schnabel, Clifford H. Spiegelman

13a. TYPE OF REPORT
Technical

13b. TIME COVERED
FROM ________ TO ________

14. DATE OF REPORT (Yr., Mo., Day)
87/8/5

15. PAGE COUNT
44

16. SUPPLEMENTARY NOTATION

17. COSATI CODES

FIELD	GROUP	SUB. GR.

18. SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
errors in variables, Monte Carlo study,
ordinary least squares, orthogonal distance regression

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Classical or ordinary least squares (OLS) is one of the most commonly used criteria for fitting data to models and for estimating parameters. This is true even when a key assumption for its use, namely that the independent variables are known exactly, is violated. Orthogonal distance regression (ODR) extends least squares data fitting to problems with independent variables that are not known exactly. Theoretical analysis, however, shows OLS is preferable to ODR for *straight line functions* under certain conditions, even when there are measurement errors in the independent variable. This has lead some to conjecture that under some similar conditions OLS will also be preferable to ODR for *nonlinear functions* even though there are errors in the independent variable.

In this paper, we present the results of an empirical study designed to examine whether ODR provides better results than OLS when there are errors in the independent variable. We examine a variety of functions, both linear and nonlinear, under a variety of experimental conditions. The results indicate that, for the data and performance criteria considered, ODR never performs appreciably worse than OLS and sometimes performs considerably better. This leads us to the conclusion that ODR is appropriate for a wide variety of practical problems.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT
UNCLASSIFIED/UNLIMITED ☒ SAME AS RPT. ☐ DTIC USERS ☐

21. ABSTRACT SECURITY CLASSIFICATION
Unclassified

22a. NAME OF RESPONSIBLE INDIVIDUAL
Dr. Jagdish Chandra

22b. TELEPHONE NUMBER (Include Area Code)
619/549-0641

22c. OFFICE SYMBOL